



Universität Regensburg

Philosophische Fakultät III
Sprach- , Literatur- und Kulturwissenschaften
Institut für Information und Medien, Sprache und Kultur (I:IMSK)
Lehrstuhl für Informationswissenschaft

Erstgutachter: Prof. Dr. Christian Wolff
Zweitgutachter: Prof. Dr. Rainer Hammwöhner

Masterarbeit

Ein Werkzeug zur datengetriebenen Visualisierung von Shakespeare-Dramen

Konzeption und Implementierung von *To See or Not to See*

Thomas Wilhelm

6. Semester M.A. Informationswissenschaft

E-Mail: Thomas.Wilhelm@stud.uni-regensburg.de

Abgegeben am 30.4.2014

Abstract

In this thesis I present the web-based and interactive text analysis tool *To See or Not to See* which visualizes 14 of Shakespeare's works in a non-linear way and offers the user data-driven visualizations. It is based on digitally annotated texts by the *Folger Digital Library* in a *TEI-conform* format. These static representations of the texts are automatically transformed to *HTML* using a *XLST*. The transformation leads to a new way to visualize otherwise linear text. Combined with standard web techniques like *JavaScript* this representation becomes an interactive tool that allows to inspect certain features of the plays like acts, scenes, characters and stage directions. There is a strong interest in quantitative aspects. Therefore graphs and diagrams are used to display the underlying data of the plays. The interest in numeric information is based on the *Distant Reading* methodology. It emphasizes the importance of data as a means to generate knowledge not only about a single but a whole group of texts. Building digital tools in order to find new ways of doing research has always been a core activity in the *Digital Humanities*. The challenges and difficulties of building a tool like *To See or Not to See* for *Distant Reading* are explored later in this thesis.

Inhalt

1	Motivation und Ziel.....	1
2	Computerunterstützte, quantitative Arbeit mit Texten	3
2.1	Digital Humanities.....	4
2.2	Distant Reading.....	10
3	Verwandte Projekte	15
3.1	Mandala Browser.....	16
3.2	Voyant Tools.....	17
3.3	WordSeer.....	18
3.4	Open Source Shakespeare.....	19
3.5	The Shakespeare Quartos Archive	19
3.6	Simulated Environment for Theatre	22
3.7	Abgrenzung von To See or Not to See.....	23
4	Konzeption und Implementierung von „To See or Not to See“	25
4.1	Zeitachse.....	26
4.2	Personenachse.....	26
4.3	Inhalt.....	27
4.4	Bühnenanweisungen.....	27
4.5	Visualisierung quantitativer Daten	28
4.6	Anwendungsszenarien und Funktionen	31
4.6.1	Informationen über Charaktere abrufen	31
4.6.2	Informationen über Akte und Szenen abrufen	36
4.6.3	Informationen zu Sprechakten abrufen	40
4.6.4	Bühnenanweisungen lesen.....	42
4.6.5	Diagramme, Graphen und Rohdaten herunterladen	44
4.7	To See or Not to See: Ein Rich-Prospect Browser?.....	45
4.8	Technischer Aufbau.....	47
4.8.1	Datengrundlage Folger Digital Texts.....	51
4.8.2	Transformationsprozess	56
4.8.3	Interaktive Webanwendung.....	59
4.9	Projektevolution.....	63
4.9.1	Bis zur Erstveröffentlichung im Mai 2013	63
4.9.2	Auswertung quantitativer Daten und Umstellung auf neues Datenmodell....	64
5	Diskussion	68
5.1	To See or Not to See: born digital?	68
5.2	Eine Anwendung des Distant Reading?.....	68
5.3	Bekannte technische Probleme.....	70
5.4	Ausblick.....	72
	Literaturverzeichnis	73

Abbildungen

Abbildung 1: Personengraph zu <i>Hamlet</i> (Moretti, 2013, S. 213).....	13
Abbildung 2: Benutzeroberfläche der Voyant Tools	18
Abbildung 3: <i>Shakespeare in quarto</i> . Vergleich zweier <i>Hamlet</i> -Drucke.....	21
Abbildung 4: Simulated Environment for Theatre (Roberts-Smith, et al., 2013)	22
Abbildung 5: Überblicksansicht und Detailfenster (unten mittig, unten rechts).....	25
Abbildung 8: Detailinformationen zum Charakter <i>Ophelia</i> in <i>Hamlet</i>	31
Abbildung 9: Gesang und Zitat in Sprechakten <i>Ophelias</i>	32
Abbildung 10: weiterführende Informationen und quantitative Daten zum Charakter <i>Ophelia</i>	32
Abbildung 11: Personennetz des Charakters <i>Ophelia</i>	33
Abbildung 12: Diagramme zu den gesprochenen Zeilen des Charakters <i>Ophelia</i>	35
Abbildung 13: Diagramme zu den gesprochenen Wörtern des Charakters <i>Ophelia</i>	36
Abbildung 14: Ausschnitt aller Sprechakte in Akt 2 in <i>Hamlet</i>	37
Abbildung 15: Personennetz für Akt 2 in <i>Hamlet</i>	38
Abbildung 16: Diagramme zur Anzahl der Zeilen, Wörter und Bühnenanweisungen in Akt2, Szene 1 von <i>Hamlet</i>	39
Abbildung 17: Verteilung der Zeilen auf die aktiven Personen in Akt 2, Szene 1 in <i>Hamlet</i> (analog für Wörter).....	40
Abbildung 18: Anzeige eines einzelnen Sprechakts am Beispiel von <i>Hamlets</i> „To be or not to be“	41
Abbildung 19: Länge von <i>Hamlets</i> Sprechakt „To be or not to be“ im Kontext.....	42
Abbildung 18: Bühnenanweisung mit Personenreferenzen	44
Abbildung 20: technischer Aufbau und Datenverarbeitungsprozess.....	48
Abbildung 21: Komponenten der Webanwendung	49
Abbildung 21: Bühnenanweisung mit Personenreferenzen im <i>TEI</i> -Ausgangsdokument	56
Abbildung 22: Vergleich der ersten (oben) und zweiten (unten) Version von <i>To See or Not to See</i>	65

Tabellen

Tabelle 1: Faktoren für <i>Rich-Prospect Browser</i> und ihre Erfüllung in <i>To See or Not to See</i> ..	45
Tabelle 2: Bei den <i>Folger Digital Texts</i> eingesetztes <i>TEI-Markup</i> und Anwendung in <i>To See or Not to See</i>	55
Tabelle 3: Schema der Relation für Charaktere mit Beispieldaten	58
Tabelle 4: Übersicht Entwicklungsphasen und Komponenten von <i>To See or Not to See</i>	67

Digitaler Anhang

Auf der beigelegten CD befindet sich der digitale Anhang dieser Arbeit. Der Inhalt der ist wie in der folgenden Tabelle dargestellt organisiert.

Inhalt	Pfad
Shakespeare-Texte von Folger in TEI-XML	\XML\
Transformationsanweisungen	\viz.xml
Transformationsengine <i>saxon</i>	\saxon\saxon9he.jar
Skript zum Starten der Transformation*	\transform.bat
Transformationsergebnisse	\htdocs\output\
CSS	\htdocs\
JavaScripts (inkl. Kopien der Bibliotheken)	\htdocs\js\
Rohdaten zum Download	\htdocs\data\
Bilddateien	\htdocs\img\
Skript für SVG -> PNG Konvertierung	\download.pl

* Um die Transformation zu starten, müssen die Pfade in der Datei **transform.bat** angepasst werden.

Nothing will come of nothing

King Lear

Akt 1, Szene 1, Zeile 99

1 Zeile: 0,29% der Szene; 0,10% des Akts; 0,75% v. Lear in Szene 1; 0,35% v. Lear in Akt 1

5 Wörter: 0,19% der Szene; 0,06% des Akts; 0,54% v. Lear in Szene 1; 0,26% v. Lear in Akt 1

Dialog mit: Cordelia, Lear's youngest daughter

Bedeutung: [*fatal error: missing NLP module*]

1 Motivation und Ziel

Unter dem Namen *Folger Digital Texts*¹ veröffentlichte die *Folger Shakespeare Library*² im Januar 2013 eine Reihe von Stücken in einer bisher nicht verfügbaren Form. Die Texte, darunter einige der wichtigsten Dramen Shakespeares, wurden als XML-annotierte Dateien zum Download bereitgestellt: „We are delighted to share our encoded texts at no cost for noncommercial uses“ (Folger Shakespeare Library, 2013a)³. Die Bibliothek stellt an derselben Stelle klar, dass ihr Ziel dabei ein moderner Umgang mit dem durch die Annotation entstandenen Potential ist: So sparen Entwickler bei der Programmierung von *Apps* und in anderen digitalen Projekten Zeit, wenn Sie diese Texte als Basis benutzen. Gegenüber den Herausgebern anderer kommerzieller Shakespeare-Editionen verschafft sich Folger damit einen Vorsprung, indem sie den *Digital Humanities* ein entscheidendes Angebot machen. Der Zugang zu einer vertrauenswürdigen, wissenschaftlich aufbereiteten Quelle ist für die Arbeit in dem jungen Feld von entscheidender Bedeutung.

Dabei sind die *Folger Digital Texts* nicht die erste und einzige digitale Ressource für die Werke des wohl am besten erforschten, englischsprachigen Autors. Verglichen mit anderen Projekten, wie z.B. *Open Source Shakespeare*, hat die TEI-konforme Kodierung jedoch einige entscheidende Vorteile. Die Texte sind so detailliert aufbereitet, dass sich neue Möglichkeiten für die Textanalyse ergeben. Diese bestehen in einer automatischen Aufbereitung, Verarbeitung und Darstellung quantitativer Aspekte des Texts. Gerade aufgrund der hohen Popularität William Shakespeares in der Wissenschaft eignen sich dessen Texte zur Untersuchung der Potentiale neuer, datengetriebener Untersuchungswerkzeuge.

Deren Erkenntnisinteresse kann über den Einzeltext hinausgehen. Die damit gewonnenen Ergebnisse können zumindest im Falle Shakespeares mit bestehenden, traditionell erarbeiteten Erkenntnissen verglichen werden. Somit können sie einen Beitrag dazu

¹ Die *Folger Digital Texts* im Web: <http://www.folgerdigitaltexts.com>, letzter Zugriff am 10.4.2014

² Die *Folger Shakespeare Library* unter Schirmherrschaft des *Amherst College* in Washington D.C. im Web: <http://www.folger.edu>, letzter Zugriff am 10.4.2014

³ Abgerufen von *Download the Code* am 28.4.2014 von <http://www.folgerdigitaltexts.org/>

leisten, automatische Verfahren gewinnbringend auf ein größeres Textkorpus auszuweiten. Schließlich sollen Ergebnisse erzielt werden, die Aussagen über eine ganze Gruppe von Texten machen. Dieser Gedanke, mittels computergestützter Methoden eine möglichst große Anzahl von Texten zu untersuchen, ist die zentrale Forderung des *Distant Reading*. Darunter sind mitunter kontroverse Ideen zu einer Neuausrichtung der literaturwissenschaftlichen Forschung gesammelt, deren auffälligster Vertreter Franco Moretti ist. Diese betreffen Methoden, Ziele und Selbstverständnis der literaturwissenschaftlichen Forschung.

Neben z.B. epistemischen Fragen, ist ein Grundproblem des *Distant Reading* seine praktische Umsetzbarkeit. Schließlich steht und fällt jede Forschungsanstrengung mit der Verfügbarkeit einer (relativ) großen Anzahl geeigneter Texte. Die automatische Verarbeitung von literarischen Texten, sofern diese z.B. über das bloße Zählen von Wörtern hinausgehen soll, braucht entsprechend annotierte Texte als Datengrundlage. Die digitale Aufbereitung von Schriftstücken ist die Lücke zwischen der Masse an Texten und den technischen Möglichkeiten. Folger hat mit den bereitgestellten, digitalen Versionen der Shakespeare-Dramen einen Beitrag zum Schließen dieser Lücke geleistet.

Das Projekt *To See or Not To See*⁴ nimmt die Folger-Texte deshalb als Möglichkeit wahr, auszuloten welche Potentiale in der automatischen Aufbereitung und Visualisierung von Dramen liegen. Es leistet somit einen Beitrag zur Diskussion um die Digitalisierung literaturwissenschaftlicher Arbeit, die durch Moretti (2007) und Ramsay (2011) befeuert wird. Bevor das Werkzeug ausführlich thematisiert wird, soll sein Kontext eingeführt werden. Diesen bilden auf oberster Ebene die *Digital Humanities* (Abschnitt 2.1). Den direkten theoretischen Rahmen bildet das bereits genannte *Distant Reading* (Abschnitt 2.2). Aus Sicht des Interfacedesigns ist *Rich-Prospect Browsing* (Abschnitt 3) für die vorgestellte Anwendung maßgeblich. Verwandte Projekte werden anschließend im selben Abschnitt vorgestellt.

⁴ Im Web unter <http://www.thomaswilhelm.eu/shakespeare/> zu erreichen.

2 Computerunterstützte, quantitative Arbeit mit Texten

McLuhan breitet in *Understanding Media* (2001) seine Theorie *The Medium is the Message* anhand vielfältiger Beispiele aus. Er argumentiert, dass jedes Medium, unabhängig von seinem Inhalt, seinen kulturellen Kontext zwangsweise verändert. Diese Veränderung sei, so McLuhan, von den Bedingungen des Mediums abhängig. Tritt demnach eine neue Art Medium auf, so wird es spezifische irreversible Veränderungen hervorrufen (S. 7 ff).

Betrachtet man den „Methodenstreit“ (Serlen, 2010; Ross, 2014), der um den Theoriekomplex *Distant Reading* in der Literaturwissenschaft entbrannt ist, kann man McLuhans Gedankengang exemplarisch beobachten. Der Computer ist als Medium und Werkzeug in die Sphäre der (Geistes-)Wissenschaft vorgedrungen. Mit seiner Verbreitung verändert sich die Weise, auf welche Menschen ihrer Arbeit nachgehen und sogar der Gegenstand ihrer Arbeit selbst. Arbeitet ein Literaturwissenschaftler mit einem Text, sind die Methoden und Ergebnisse dann abhängig vom Medium in dem der Text repräsentiert ist? Es scheint, dass diese Frage den Kern der Debatte um die computerunterstützte, datengetriebene Textanalyse bildet: Welche Veränderungen sind epistemologisch zu erwarten, welche Methoden sind davon betroffen oder entstehen neu und wie wird sich das Bild der Disziplin nach innen wie nach außen entwickeln? Unter anderem Ramsay (2011) widmet sich diesen Fragen intensiv. Wie auch immer sich eine datengetriebene, quantitative Textanalyse etablieren sollte – die Literaturwissenschaft als Teildisziplin der Geisteswissenschaften kann davon nicht unberührt bleiben (Schreibman, Siemens, & Unsworth, 2004).

Die Tatsache, dass unvermeidbare und fundamentale Veränderungen auf diesen Wissenschaftsbereich zukommen bzw. im Gange sind, kann ein Anhaltspunkt für die Heftigkeit der Debatte sein. Bisweilen sind es die Akteure selbst, die durch die Vehemenz, mit welcher sie ihre Argumente vorbringen, ihren Beitrag (im doppelten Sinne) dazu leisten.⁵ Einer dieser Akteure, eine zentrale Figur in der Debatte um die Zukunft der Literaturwissenschaft, ist dabei Franco Moretti. Auf ihn geht die Idee des *Distant Reading* zurück, die mittlerweile eine Reihe von Unterstützern gefunden hat und seit 2010 im *Stanford Literary Lab* institutionalisiert ist (Stanford Literary Lab).

⁵ Ein Indiz hierfür ist, dass Moretti die Veröffentlichungen seines *Literary Lab* in Stanford als „Pamphlete“ bezeichnet.

(Burdick, Drucker, Lunenfeld, Presner, & Schnapp, 2012) versuchen die Debatte um die Digitalisierung der Geisteswissenschaften im Allgemeinen zu entschärfen:

Digital Humanities is an extension of traditional knowledge skills and methods, not a replacement for them. Its distinctive contributions do not obliterate the insights of the past, but add and supplement the humanities' long-standing commitment to scholarly interpretation, informed research, structured argument, and dialogue within communities of practice. (S. 16)

Texte aus der Distanz – also beinahe unabhängig von ihrem Inhalt und anhand ihrer quantitativen Daten – „lesbar“ zu machen ist eine der Herausforderungen dieses Projekts. Dabei wird den Argumenten des *Distant Reading* praktisch auf den Grund gegangen. Welche davon können nachvollzogen werden? Welche Probleme bestehen oder treten erst mit Entwicklung des Werkzeugs im Hinblick auf die Theorie auf? Um diesen Fragen nachgehen zu können, folgt nun eine systematische Auseinandersetzung mit dem theoretischen Rahmen des Projekts, den *Digital Humanities* (Abschnitt 2.1) und dem *Distant Reading* (Abschnitt 2.2).

2.1 Digital Humanities

Seitdem das Digitale in die Welt der Geisteswissenschaften vorgedrungen ist, wird versucht das Phänomen einzugrenzen und zu definieren. Zuerst als *Humanities Computing* oder *Computing in the Humanities* bezeichnet, verschiebt sich die Bedeutung später in Richtung der *Humanities*, die computerunterstützt arbeiten, also begrifflich zu *Digital Humanities* (Berry, 2011, S. 3). (Svensson, 2009) benennt diesen Wandel und erklärt ihn mit einer Veränderung der Forschungsgegenstände von Digitalisierten zum genuin Digitalen.

Während es in der Frühphase hauptsächlich die Informatikkomponente war, die Aufwand, Möglichkeiten und Fragestellungen bedingte, so hat sich das Portfolio heute weit darüber hinaus entwickelt. Am Anfang lag der Fokus auf der Digitalisierung bestehender Ressourcen, hauptsächlich von Texten. Dieser wurde mit der Zeit in Richtung Aufbereitung der digitalisierten Texte und Multimedia neu gefasst (Schreibman et al., 2004).

Aus Sicht der Forschungsobjekte ist die Digitalisierung abgeschlossen, da sowohl digitalisiertes Material wie auch Ressourcen verwendet werden, die von vorne herein

digital sind. Obwohl Forschungsgegenstände heute analoge wie digitale Artefakte umfassen, ist die Digitalisierung von analogen Kulturgütern jedoch weder vollständig noch sind sämtliche ihrer Möglichkeiten erforscht. Die Grenzen, welche der Forschung durch die technischen Möglichkeiten gesetzt werden, bestehen dabei weiter, auch wenn die relativen Kosten seit den Anfängen der *Digital Humanities* enorm zurückgegangen sind (Busa, 2004).

Für die Arbeit mit digitalen Forschungsobjekten, ob von vorne herein digital (*born digital*) oder nicht, können computerbasierte Werkzeuge verwendet werden. Beschränkte sich die automatische Verarbeitung von Texten zuerst auf Aufgaben wie das Zählen von Worthäufigkeiten, stehen mittlerweile ungleich komplexere Analysemöglichkeiten zur Verfügung. Die Weiterentwicklung von Werkzeugen ist dabei eng mit der ihrer Ressourcen verknüpft. Um Texte „aus der Distanz lesbar“ zu machen, also automatisiert Informationen gewinnen zu können, braucht es neben der reinen digitalen Erfassung eines Texts eine Anreicherung mit zusätzlichen Informationen (Crane, 2006). Erst durch die Annotation – geschieht diese manuell oder wiederum automatisiert – also das Hinzufügen von Information zu bestehenden Daten wird eine Verarbeitung durch Software möglich bzw. vereinfacht (Garside, Leech, & McEnery, 1997, S. 4 f.). Garside et al. (1997) weisen zudem auf eine erhebliche Aufgabe hin, indem sie Anstrengungen zur Standardisierung der Annotation fordern (S. 6 f.). Mit den Vorgaben der *Text Encoding Initiative* (TEI), auf denen die Ausgangstexte für *To See or Not to See* basieren, ist diese Forderung verwirklicht.

Über Gegenstände und Werkzeuge hinaus sind die digitalen Geisteswissenschaften nach Burdick et al. (2012) heute eher Bedingungen unter denen Forschung betrieben und veröffentlicht wird, als ein eigenständiger Wissenschaftszweig:

[...] the advent of Digital Humanities implies a reinterpretation of the humanities as a generative enterprise: one in which students and faculty alike are *making things* as they study and perform research, generating not just texts (in the form of analysis, commentary, narration, critique) but also images, interactions, cross-media corpora, software, and platforms. (S. 10)

Die Autoren sprechen von einer Neuinterpretation der Geisteswissenschaften als produktive Disziplin, deren Ergebnisse in multimedialer Form dargereicht werden. Hierin schwingen wieder die Erstellung digitaler Ressourcen und Werkzeuge mit. Deutlich wird zudem eine Veränderung des Veröffentlichungsprozesses hin zu multimedialen Produkten. Traditionelle Veröffentlichungen in ausschließlich schriftlicher Form werden hierdurch und durch Interdisziplinarität in Frage gestellt. In vielen Fällen, gerade bei größeren Vorhaben, leisten vergleichsweise viele Personen aus unterschiedlichen Disziplinen jeweils ihren Beitrag (Burdick et al., 2012, S. 84). Eine weitere Bedingung unter denen in den *Digital Humanities* geforscht wird, ist laut Burdick et al. (2012) eine Kultur des Ausprobierens, in der Risiko und Scheitern als integraler, positiver Bestandteil des Produktionsprozesses verstanden wird (S. 21 f.).

(Svensson, 2010) sieht die *Digital Humanities* ebenfalls nicht als eigenständiges Feld. Mithilfe eines analytischen Modells identifiziert er fünf Dimensionen, in denen Berührungspunkte zwischen den Geisteswissenschaften und dem Digitalen bestehen. Zu Svenssons *modes of engagement* gehören, wie oben, Veröffentlichungsstrategien, bei denen z.B. digitale Artefakte entstehen. So wie Burdick et al. das Scheitern, also z.B. die schrittweise Entwicklung von Software, als Teil der Kultur der *Digital Humanities* sehen, abstrahiert Svensson weiter und sieht das Feld als „experimentelles Labor“. Vorstöße wie sie Moretti mit *Distant Reading* vornimmt, benutzen die es, folgt man Svensson, als *activist venue*. Die letzten beiden Dimensionen sind laut Svensson die zwei wichtigsten bei der Arbeit in den *Digital Humanities* und wurden entsprechend schon weiter oben thematisiert: Die Informationstechnologie als Werkzeug und digitale Forschungsgegenstände. Er legt dabei Wert darauf, dass jede dieser Dimensionen weder trennscharf noch exklusiv ist. Die fruchtbaren Wechselwirkungen zwischen digitalen Ressourcen und Werkzeugen wurden eingangs bereits dargestellt.

Des Weiteren schreibt Svensson (2010, Para. 34 f.) Bibliotheken eine besondere Rolle in den *Digital Humanities* zu. Er stellt eine Analogie zwischen Bibliothek und Labor her, die durch das kollaborative Schaffen von Wissen und die Archivfunktion bestimmt ist. Die Herausforderungen, Chancen und Folgen der Digitalisierung für Bibliotheken thematisiert Crane (2006) und sieht diese in sechs Bereichen: Die Größe, die digitale Bibliotheken annehmen werden, übersteigen bisherige Inventare bei weitem. Mit der Größe

wird auch das inhaltliche Spektrum breiter, weil die Kollektionen um zusätzliche Themen und Sprachen erweitert werden. Die Genauigkeit mit der die Texte ausgezeichnet werden bestimmt in Zukunft die Möglichkeiten des Informationsabrufs und der Wissensbildung. Die Häufigkeit von Fehlern, wie sie z.B. in der Digitalisierung von Texten auftreten, ist an das Wachstum der Bibliotheken gebunden. Zuletzt liegen laut Crane im Publikum und der Reichweite von Distributoren rechtliche, ökonomische und nicht zuletzt philosophische Unwägbarkeiten. Mit den *Digital Texts* hat Folger sich einigen dieser Herausforderungen gestellt, indem sie Ressourcen bereitgestellt haben, um ihre Sammlung zu digitalisieren. Daran ist unter anderem die Wechselwirkung zwischen Zeitaufwand und Annotationstiefe gut zu beobachten. Zum jetzigen Zeitpunkt sind noch nicht sämtliche Dramen Shakespeares im *TEI*-Format veröffentlicht. Für eine Bibliothek von den Ausmaßen, wie sie Crane aufzeichnet, kann die manuelle Annotation demnach keine effiziente Arbeitsweise darstellen. Dies wirft die Frage auf, ob eine derart genaue Auszeichnung überhaupt für alle Texte machbar bzw. notwendig ist.

Welche Möglichkeiten sich jedoch aus der reinen Verfügbarkeit von Millionen digitalen Büchern ergeben, wird im Folgenden an einem Beispielprojekt deutlich werden. Auf dem Weg zu den *Digital Humanities*, wie sie gerade vorgestellt wurden, liegen eben solche wegweisenden Projekte. An diesen Meilensteinen soll nun die Entwicklung ausgehend vom *Humanities Computing* exemplarisch nachvollzogen und Kerneigenschaften identifiziert werden, um dem Wesen der *Digital Humanities* auf den Grund zu gehen.

Den Ausgangspunkt bildet 1949 der *Index Thomisticus* von Roberto Busa. Mit Hilfe von IBM erstellte der katholische Geistliche die erste, digitale Konkordanz der gesammelten Werke von Thomas von Aquin. Das Vorhaben konnte mit traditionellen Methoden nicht verwirklicht werden, es verlangte nach technischer Unterstützung, so Busa (1980): „It was clear to me, however, that to process texts containing more than ten million words, I had to look for some type of machinery“ (S. 83). Das definierende Kriterium für das Digitale an geisteswissenschaftlicher Arbeit ist hier, zuerst in Form von Lochkarten (ab 1949) und wenige Jahre später Magnetbändern (und schließlich einer modernen

Datenbank⁶), zum ersten Mal erfüllt: Aus einer analogen Quelle wird eine digitale Resource und Anwendung geschaffen.

Busas Projekt zeichnet nicht nur deshalb den Weg für die *Digital Humanities* vor. Die Erstellung des *Index Thomisticus* war genuin interdisziplinär: Durch die Beteiligung der IBM und Busa arbeiten Informatik und Sprachwissenschaft eng zusammen und bringen so die Computerlinguistik mit auf den Weg (Busa, *The Annals of Humanities Computing: The Index Thomisticus*, 1980, S. 86). Auf globaler Ebene arbeitet eine große Gruppe von Personen mit unterschiedlicher Ausbildung zusammen. Zwar ist die Zusammenarbeit unterschiedlicher Fachbereiche oder das Einbinden externer Partner in den Hochschulen kein neues Konzept, scheint aber insbesondere in den Geisteswissenschaften neu aufzuleben. Diese Art von Interdisziplinarität und die damit einhergehende Offenheit führt laut Burdick et al. (2012, S. 25 f.) zu einer verbesserten Wahrnehmung der Relevanz geisteswissenschaftlicher Forschung in der Öffentlichkeit.

Während Busa mit einem einzigen Autor arbeitet, um dessen Sprache genauer zu verstehen, beginnt 1991 in Großbritannien die Arbeit am *British National Corpus*⁷ durch ein Konsortium aus Universitäten, Herausgebern von Wörterbüchern und Bibliotheken. Es umfasst seit seiner Fertigstellung im Jahr 1994 100 Millionen Wörter sowohl geschriebener wie auch gesprochener Sprache aus unterschiedlichsten Quellen. Die aktuellste Version von 2007 (*BNC XML Edition*) ist zum *TEI*-Format kompatibel. Nach Auswahl geeigneter Texte – das Korpus sollte möglichst breit angelegt sein – wurde die Digitalisierung und Annotation zum großen Teil automatisch vorgenommen (British National Corpus, 2014). Das *BNC* ist ein Abbild der englischen Sprache am Ende des 20. Jahrhunderts. Hier wird eine enorme Datenmenge auf einfache Art für eine große Personengruppe zugänglich, ein Mehrwert den man auch in Deutschland erkannt hat. Denn ein ähnlicher Versuch die gegenwärtige geschriebene deutsche Sprache (und mittlerweile auch andere) zu erfassen ist der *Leipziger Wortschatz*⁸, der hauptsächlich automatisiert aus digitalen Quellen erhoben wird (Leipzig Corpora Collection, 2014).

⁶ Eine Onlineversion des Index ist seit 2005 unter <http://www.corpusthomisticum.org/> (letzter Zugriff am 9.4.2014) zu erreichen. Das Projekt erstreckt sich somit über einen Zeitraum von mehr als 50 Jahren.

⁷ Im Web unter <http://www.natcorp.ox.ac.uk/>, letzter Zugriff am 15.4.2014

⁸ Im Web unter <http://wortschatz.uni-leipzig.de/>, letzter Zugriff am 15.4.2014

Während das *BNC* eine Momentaufnahme darstellt, geht das *Culturomics* Projekt von (Michel, et al., 2011) weiter, indem es sich das Ziel setzt auch sprachliche Entwicklungen nachvollziehbar zu machen. Zu diesem Zweck wurden Bücher aus den Jahren zwischen 1800 und 2000 digital erfasst. Somit kann das Projekt exemplarisch für das Wesen und den aktuellen Stand der *Digital Humanities* stehen, da es viele der Kernaspekte in sich vereint. Interdisziplinarität, das Überwinden von Einrichtungsgrenzen, Veröffentlichungspraxen und Methodik (die Arbeit mit enormen Datenmengen und die Anwendung geisteswissenschaftlicher Methoden zur Interpretation). Hauptsächlich beteiligt sind verschiedene Einrichtungen der Harvard University, Google und weltweit über 40 Bibliotheken (Michel, et al., 2011, S. 176). Die Nennung des gesamten *Google Books Teams* als ein einzelner Mitautor, veranschaulicht wie die Forschungspraxis bei einem derartigen Großprojekt mit der Veröffentlichung in Form eines Zeitschriftenartikels haddert.⁹ Die Forschungsergebnisse des Teams können über einen alternativen Veröffentlichungskanal nachvollzogen werden, indem weiteres Material und das benutzte Werkzeug online¹⁰ bereitgestellt werden. Dem produktiven und multimedialen Wesen der *Digital Humanities* wird deshalb hier – und auch im vorliegenden Projekt – Rechnung getragen. Entscheidend ist, dass Ergebnisse wie jene zur Zensur im Nationalsozialismus (Michel, et al., 2011, S. 181), nicht allein durch den Einsatz einer digitalen Ressource und eines digitalen Werkzeugs, dem *Ngram Viewer*, zustande kommen. Dieses liefert quantitative Daten deren Erklärung und Fruchtbarmachung anderer Methoden bedarf. Externe Quellen für Erklärungsmodelle anzuzapfen, wie im obigen Beispiel zu Zensurvorgängen etwa die Geschichtswissenschaft, ist beim *Distant Reading* ein zentrales Motiv, wie der kommende Abschnitt 2.2 zeigen wird.

Anhand der drei Beispiele wurde deutlich, wie sich Projektansprüche und -umfänge im Lauf der Zeit verändert haben. Für dieses Wachstum gibt es viele Gründe. Einer davon ist die fortschreitende Institutionalisierung der *Digital Humanities*. Die erste Fachzeitschrift war 1966 *Computers and the Humanities* (Burdick et al., 2012, S. 123). Während

⁹ Derart lange Autorenlisten sind für Großprojekte zwar nicht unüblich, aber im alltäglichen Wissenschaftsbetrieb eher die Ausnahme. Die Nennung eines Teams ist in diesem Zusammenhang auffällig.

¹⁰ Im Web sind unter <http://www.culturomics.org/> das Werkzeug *Google Books Ngram Viewer*, die Forschungsrohdaten, der Artikel und das zugehörige Twitterprofil zu finden. Letzteres ist einerseits Marketingstrategie und andererseits eine Aufforderung zur Auseinandersetzung und Kritik durch die Öffentlichkeit.

in den USA mit der *Association for Literary and Linguistic Computing* und der *Association for Computers and the Humanities* in den 1970er Jahren erste Einrichtungen geschaffen werden (Burdick et al., 2012, S. 123), dauert es zur Gründung der Gesellschaft *Digital Humanities im deutschsprachigen Raum* bis ins Jahr 2012 (DHd, 2014). Im März 2014 fand ihre erste Jahrestagung an der Universität Passau statt.

Diese kurze Einführung hat gezeigt, mit welchen Eigenschaften die *Digital Humanities* beschreiben werden können. Im Folgenden wird aus ihren vielen Disziplinen die Literaturwissenschaft herausgegriffen und mit dem *Distant Reading* eine durch das Digitale motivierte Anwendung vorgestellt. Für diese ist die sowohl die computerunterstützte Arbeit mit Daten zu literarischen Werken wichtig, also eher *Humanities Computing*, wie auch interdisziplinäre Arbeit zum Zwecke der Theoriebildung und die Visualisierung quantitativer Daten, demnach eher *Digital Humanities*.

2.2 Distant Reading

Distant Reading ist einerseits der Begriff, der sich für die Ideen Morettis seit ihrer ersten Veröffentlichung herauskristallisiert hat. Andererseits ist es ein Reizwort für die Literaturwissenschaften und andere, betroffene oder vereinnahmte Felder. Das gleichnamige Buch Morettis (2013) und zuvor *Graphs, Maps, Trees* (2007) sind Sammlungen verschiedener Aufsätze zum Thema. In dieser fragmentarischen Form wird dabei immer wieder der Kern wiederholt: Alleine die Anzahl verfügbarer Schriften erfordert eine Neuausrichtung literaturwissenschaftlicher Arbeit, sowohl in der Methodik wie auch in den Zielen, um das darin steckende Potential nutzbar zu machen.

Motiviert ist Morettis Ansatz durch die Kluft zwischen der Anzahl an Texten, die der literaturwissenschaftlichen Forschung zur Verfügung stehen¹¹ einerseits und der Beschäftigung der Forschung mit einem sehr kleinen Teil dieser Texte¹² andererseits. Moretti will den literaturwissenschaftlichen Kanon nicht verändern, sondern aus ihm ausbrechen (Moretti, 2013, S. 66) und hält ihn laut Ross (2014, S. 3) sogar für elitär. An Morettis Unzufriedenheit mit der Definition von Genres beobachtet Serlen (2010, S. 216)

¹¹ Im englischen als *archive* bezeichnet. Damit sind sämtliche jemals publizierten literarischen Texte gemeint.

¹² Moretti gebraucht hierfür den Begriff „Kanon“, den er nicht weiter bestimmt. Damit ist eine Gruppe von Texten gemeint, die in der Literaturwissenschaft besondere Aufmerksamkeit erhalten, deren Ränder aber unscharf und veränderbar sind.

diese Motivation. Von einem Text aus (s)ein Genre zu definieren, bringt ähnliche Texte in die Situation, am Unterschied zur „Normalform“ gemessen zu werden. Dieser intentionalen Form, die literarische Landschaft zu erklären, hält Moretti eine eher extensionale Vorgehensweise entgegen. Aus *Close Reading* soll *Distant Reading* werden, das Zusammenhänge und Makrostrukturen auf andere, vermeintlich adäquatere Weise erklären kann.

Mit dem „Lesen aus der Distanz“ ist eigentlich das „Nicht-Lesen“ eines einzelnen Texts, also die Abstraktion einzelner Teilaspekte vieler Texte gemeint, um dadurch globale Phänomene erst sichtbar und erklärbar zu machen. Burdick et al. (2012) bringen es auf den Punkt:

[...] distant reading explicitly ignores the specific features of any individual text that close reading concentrates on in favor of gleaning larger trends and patterns from a corpus of texts. Distant reading is therefore not just a “digitization” or “quickener” of classic humanities methodologies. It is, rather, a new way of doing research wherein computational methods allow for novel sets of questions to be posed about the history of ideas, language use, cultural values and their dissemination, and the processes by which culture is made. (S. 39)

Die Aufgabe besteht darin, Schlüsselaspekte in einer Gruppe von Texten zu identifizieren und auszuwerten. Auf Basis dieser Daten soll schließlich eine Erklärung gefunden werden. Es handelt sich also um einen zweistufigen Prozess: Zuerst müssen Daten gesammelt und anschließend ein Erklärungsmodell für diese Daten gefunden werden.

Im ersten Schritt des *Distant Reading*, bei der Abstraktion des Texts und der Auswahl einer spezifischen Untersuchungsgröße sieht Serlen (2010) die Gefahr der Beliebigkeit und Unschärfe: „Moretti runs the risk of identifying genres and devices whose totality is as artificial as that of the individual texts he is trying to displace“ (S. 217). Die Aufgabe, geeignete Eigenschaften eines Textes zu identifizieren, die überhaupt eine Aussagekraft besitzen und nicht etwa nur zufällig zu einem anderswo beobachteten Phänomen (z.B. Verkaufszahlen eines Buchs, s.u.) passen, ist dabei ein erhebliches Problem. Man kann argumentieren, dass dies im Sinne der *Trial-and-Error*-Kultur in den *Digital Humanities* keinen Hinderungsgrund darstellt. Es bleibt offen, ob ein Forscher mit der Erfahrung, wie sie Moretti mitbringt, die sprichwörtliche Nadel im Heuhaufen eher finden kann oder doch den Wald vor lauter Bäumen nicht sieht. Vor allem letzteres beleuchtet eine

Argumentationsstrategie, die in den *Digital Humanities* oft, wie auch hier, bemüht wird. Die reine Sichtbarmachung bestimmter Daten oder Phänomene kann zu neuen Erkenntnissen oder Fragen führen. Die Art und Weise der Interpretation der Darstellungen stellt dabei ein eigenständiges Problem dar, das im zweiten Schritt des *Distant Reading* methodisch angegangen wird.

Die Modelle zur Erklärung dieser Visualisierungen liegen nicht innerhalb der Texte, sondern werden von Moretti importiert. Ökonomische, politische, ökologische Bedingungen und andere Disziplinen, wie z.B. die Evolutionstheorie, werden zur Theoriebildung herangezogen. Externe Kräfte (*forces*), wie Moretti sie nennt, beeinflussten die (Weiter-)Entwicklung des Systems Literatur. Welche spezifischen Formen sich durchsetzen, z.B. innerhalb eines Genres, unterliege Mechanismen wie jenen des Marktes (Moretti, 2013, S. 69 f.). Er versucht diese Theorie mit einem Beispiel zu untermauern. Warum setzt sich ein bestimmter Krimiautor gegen andere durch? Hierzu extrahiert er aus einer großen Zahl von Kriminalromanen das Vorhandensein von „Hinweisen“ als die Untersuchungsgröße. Es stellt sich heraus, dass jene Texte erfolgreicher verkauft wurden, in denen Hinweise vorkommen und diese bestimmte Kriterien erfüllen (Moretti, 2013, S. 71 ff.). Mittels eines einzelnen Beobachtungspunkts in vielen Texten nähert er sich so einem Modell zur Erklärung des größeren Zusammenhangs an.

Für den beschriebenen Perspektivwechsel, der durch die Sichtbarmachung von Texteigenschaften anhand datengetriebener Darstellungen entsteht, werden Visualisierungstechniken aus anderen Disziplinen herangezogen (z.B. im Artikel *Style Inc.* bei Moretti, 2013, S. 179 ff.) oder nach neuen Formen gesucht (z.B. bei Keim & Oelke, 2007). Moretti nutzt die Vorteile von Visualisierungen für seine Beispiele. Mit *Graphs, Maps, Trees* (2007) legt er die Grundsteine für eine auf der Visualisierung quantitativer Daten beruhende Literaturtheorie. Mit dem Artikel *Network Theory, Plot Analysis*, hier aus *Distant Reading* (2013, S. 211 ff.), geht er den nächsten Schritt, indem er aus einem Personengraphen auf den Inhalt von Shakespeares *Hamlet* schließt (Abbildung 1).

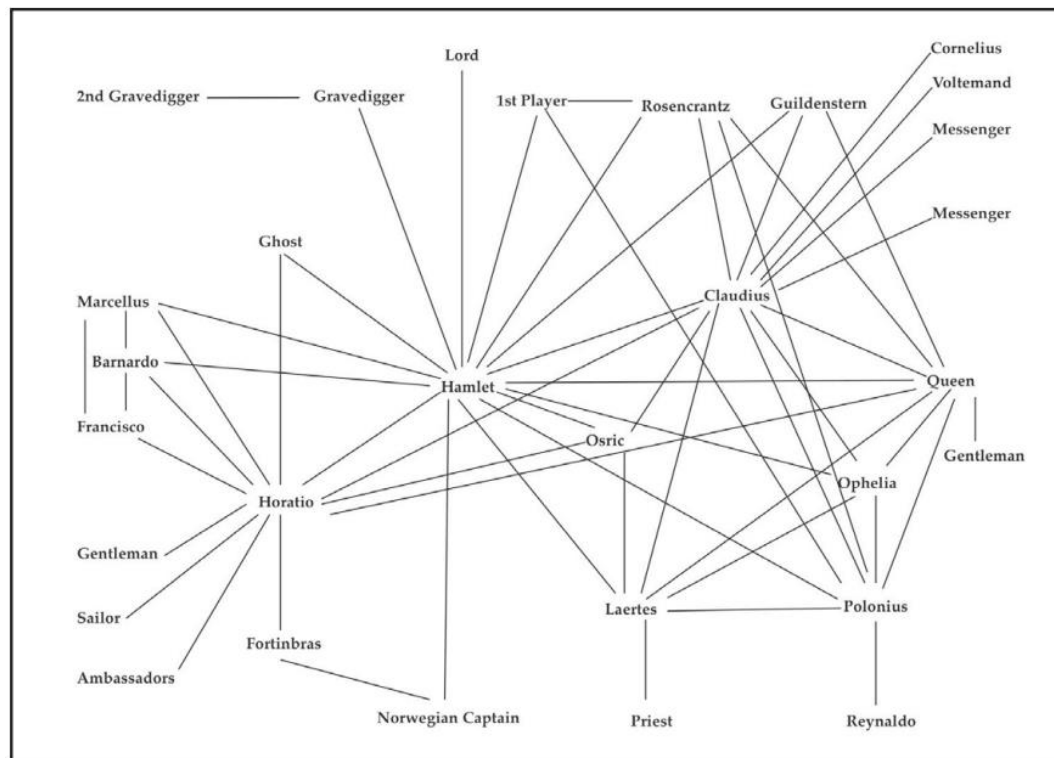


Abbildung 1: Personengraph zu *Hamlet* (Moretti, 2013, S. 213)

Durch die netzartige Darstellung der Beziehungen zwischen den einzelnen Personen wird die Zeit vom Kontinuum zum Moment. Diese Reduktion hat zur Folge, dass Zusammenhänge zwischen Charakteren sichtbar werden und hieraus Gruppen abgeleitet werden können, für die bestimmte Eigenschaften zutreffen. Beispielsweise wählt Moretti (2013, S. 217) alle Knoten aus, die sowohl mit *Claudius* wie auch *Hamlet* verbunden sind, um ein „tödliches Subnetz“ sichtbar zu machen. Beinahe alle Charaktere, die diese Eigenschaft erfüllen, sterben einen gewaltsamen Tod. Die individuelle Täterschaft ist für Moretti dabei nicht mehr erheblich, vielmehr entscheidet die Position im Graphen, also das Umfeld der Charaktere, über deren Schicksal. In weiteren Beispielen entfernt er bestimmte Knoten, um die Bedeutung dieser für das Stück zu untersuchen.

Zwar kann man ihm hier den Vorwurf machen, eigentlich *Close Reading* zu betreiben. Schließlich geht es nur um einen ausgewählten Text und zudem schwächen Vorwissen und die manuelle Erstellung des Netzes sein Argument. Dennoch bleibt es ein Überzeugendes, da der eigentliche Text hier nicht gelesen wird. Kennt man ein Theaterstück nicht so genau wie hier im Beispiel *Hamlet*, handelt es sich also um ein Stück au-

ßerhalb des Kanons, so kann ein Netz aus Personen eine Orientierungshilfe mit Interpretationspotential sein. Aus diesem Grund wurden in *To See or Not to See* ähnliche Graphen integriert, die automatisch und für jeden Akt einzeln erstellt werden.

Um nicht in einem relativ kleinen Korpus verhaftet zu bleiben, also außerhalb des Kanons zu agieren, braucht es für das *Distant Reading* Bibliotheken, wie Crane (2006) sie aufzeigt. Erst wenn größere Datenmengen verfügbar seien, könne das volle Potential der Methode ausgeschöpft werden, so Moretti (2013, S. 240 nach Ross, 2014, S. 4). Mit dem TEI-Standard bestehen technische Mittel, um wie bei *To See or Not to See Distant Reading* zu ermöglichen. Die praktische Umsetzbarkeit einer Annotation, wie sie für die Analyse nötig ist, wurde hinsichtlich der Menge zu digitalisierender Texte weiter oben bereits in Frage gestellt. Praktische Probleme der Methode werden weiter unten am Beispiel der vorliegenden Anwendung erörtert (Abschnitt 5.2). Nun sollen jedoch zuerst verwandte Projekte vorgestellt werden, um das Feld im dem es sich bewegt genauer auszuleuchten.

3 Verwandte Projekte

Im Folgenden werden einige Projekte vorgestellt, die sich den Methoden des *Distant Reading* annähern oder als *Rich-Prospect Browser* klassifizieren lassen. Darunter sind auch Vorhaben, die mit Inhalten von Shakespeare arbeiten. Die Präsenz dieses Autors in vielen Forschungsarbeiten kommt nicht von Ungefähr, schließlich wurde schon in den 1960er Jahren damit begonnen seine Werke zu digitalisieren (Mullin, 2003). *Voyant Tools* und *Open Source Shakespeare* wurden bereits am gleichen Ort kurz eingeführt. Im Zuge der Recherche zum Vergleich von Editionen trat schließlich das *Shakespeare Quartos Archive* in Erscheinung. Das neueste Projekt, eine 3D-Simulationssoftware für Shakespeare-Dramen, dessen Ergebnisse erst kurz vor Abschluss dieser Arbeit veröffentlicht wurden, schließt diesen Teil ab. Zuvor soll jedoch eine entscheidende Gemeinsamkeit einiger dieser Vorhaben mit *To See or Not to See* (ausführlicher als bei Wilhelm, Burghardt, & Wolff, 2013) präsentiert werden.

Diese Verwandtschaft liegt in der Art, wie Informationen dargestellt werden. Ruecker, Radzikowska, & Sinclair (2011) führen den Begriff des *Rich-Prospect Browsing* ein. Darunter verstehen sie Anwendungen, die Benutzer bei der Informationssuche ohne konkret formuliertes Ziel durch das Erforschen (*Browsing*) des Suchraums zum Erfolg verhelfen sollen (S. 2). Das Verstehen von Zusammenhängen innerhalb einer Sammlung von Daten steht an vorderster Stelle. Hierzu müssen diese Werkzeuge die folgenden Kriterien erfüllen (S. 3 f.):

- Der Startbildschirm zeigt eine Übersicht über alle Informationseinheiten, die mit dem Werkzeug bearbeitet werden können, so etwa allen Bildern einer Galerie.
- Es gibt die Möglichkeit zur Neuordnung dieser Informationseinheiten
- Jede Einheit ist Zugangspunkt zu weiteren, zugehörigen Informationen
- Die Einheiten folgen einer ihnen inhärenten Ordnung und geben die verwendbaren Werkzeuge vor
- Wenn möglich kann der Benutzer zwischen verschiedenen Darstellungsformen der gleichen Datenkollektion wählen
- Die visuelle Ordnung der Übersicht soll dem Nutzer sogleich Aufschluss über ihre Bedeutung geben
- Der Benutzer soll die einzelnen Einheiten markieren können, um sie im Falle einer Neuordnung (z.B. Sortieren) wiederzuerkennen.

Die erste Forderung der obigen Liste schafft eine grundlegende Vorstellung dessen, wie ein *Rich-Prospect Interface* aussehen kann. Eine Sammlung vieler einzelner bzw. gruppierter Elemente, mit denen der Nutzer interagieren kann (*Browser*), um einen Mehrwert zu erhalten (*Rich-Prospect*). Diese Interaktionsform ermöglicht eine visuelle Abstraktion der Anfragen an die Datenbasis, sei es z.B. an einen annotierten Text oder eine relationale Datenbank (Ruecker et al., 2011, S. 23). Somit wird dem Benutzer die Arbeit erleichtert und gleichzeitig seine Möglichkeiten eingeschränkt. Je nach Kontext und Anwender kann dies positive oder negative Folgen für die Nützlichkeit des Werkzeugs haben.

Des Weiteren sehen Ruecker et al. (2011) in der digitalen Erhaltung von Kulturerbe ein Anwendungsfeld, für das sich *Rich-Prospect Interfaces* besonders eignen (S. 24). Sie argumentieren, dass derartige Vorhaben typischerweise mit Datenkollektionen geeigneten Umfangs befasst sind. Die Anzahl der Informationseinheiten liege im Bereich von Hunderten oder Tausenden. Diese Einschränkung kann aus der Erfahrung, die mit *To See or Not to See* gemacht wurde, aber auch in der praktischen Unmöglichkeit liegen, eine noch größere Anzahl von Elementen gleichzeitig auf dem Bildschirm anzuzeigen (siehe Abschnitte 4.3 und 4.7). Daneben hätten laut Ruecker et al. (2011) kulturelle Gegenstände oft besonders geeignete Metadaten und lassen sich so für den Benutzer ersichtlich in einer Übersicht darstellen (S. 24). Diese Aussage ist zumindest für das vorliegende Projekt zutreffend.

3.1 Mandala Browser

Beim *Mandala Browser*¹³ von Sinclair et al. (2011) handelt es sich um ein Werkzeug, das dem Nutzer einen visuellen Zugang zu Texten gewährt. Auf Basis dieser Darstellung sollen schließlich Interpretationen eines oder mehrerer Texte möglich werden. Die Autoren haben mit dem Werkzeug den Schulkontext im Blick. Es ist Teil einer Reihe von Veröffentlichungen des *Experimental Reading Workshop*¹⁴ unter dem *Text Analysis Portal for Research*¹⁵ (TAPoR) der Universität Alberta.

¹³ Im Web unter <http://mandala.humviz.org/>, letzter Zugriff am 17.4.2014

¹⁴ Im Web unter <http://humviz.org/>, letzter Zugriff am 17.4.2014

¹⁵ Im Web unter <http://www.tapor.ca/>, letzter Zugriff am 18.4.2014

Die Anwendung wird von den Autoren ähnlich wie *To See or Not to See* anhand XML-annotierter Texte von Shakespeare vorgestellt, kann aber unterschiedlichste XML-Strukturen abbilden. Als Datengrundlage dient das Korpus des *WordHoard*-Projekts der Northwestern University¹⁶. Es handelt sich insofern um ein *Rich-Prospect Interface*, da es zuerst alle Informationseinheiten anzeigt, die der Benutzer auswählt. Hier besteht die Möglichkeit direkt auf Ebenen des XML-Baums Bezug zu nehmen, also z.B. alle Sprechakte als die darzustellende Einheit zu wählen. Jede Einheit wird dabei als Punkt dargestellt. Die mandalaartige Darstellung wird dann durch die Gruppierung dieser Punkte erreicht. Hierfür stehen dem Dokument eigene Aspekte zur Verfügung. Im Fall von Dramen z.B. Charaktere. Es ist möglich, weitere Eigenschaften zur Gliederung zu verwenden und so entsteht eine baumartige Struktur, bei dem Einheiten zusammenhängen, die sich bestimmte Eigenschaften teilen. Im Grunde handelt es sich bei diesem Prozess um eine Aneinanderreihung verfeinerter Anfragen an die Datenbasis, die eine visuelle Filterung zur Folge hat. Aus dem sich ergebenden Bild sollen dann, im Sinne des *Distant Reading*, Schlussfolgerungen gezogen werden können.

3.2 Voyant Tools

Mit dem Slogan *Reveal Your Texts* bewerben Rockwell, Sinclair, Ruecker, & Organisciak (2010) ihren digitalen Werkzeugkasten *Voyant*¹⁷ (früher: *Voyeur Tools*). Die enthaltenen Werkzeuge ermöglichen die Analyse der geladenen Texte auf Wortebene. Es besteht die Möglichkeit mehrere beliebige, nicht annotierte Texte zu laden oder aus Beispieltexten, so z.B. allen Shakespeare-Texten der *Moby Edition* (siehe *Open Source Shakespeare*), zu wählen.

Zu den Werkzeugen (Abbildung 2) gehören eine Übersicht (1) zu den geladenen Texten (Anzahl der Texte, Worthäufigkeiten, herausstechende Wörter, usw.) und eine *Word Cloud* (2; *Cirrus*), die häufige Wörter größer und zentraler darstellt als weniger häufige. Des Weiteren werden alle enthaltenen Wörter eines Textes in der gewählten Kollektion aufgelistet (3) und ihre (relative) Häufigkeit sowie Verteilung über die Länge der

¹⁶ Im Web unter <http://wordhoard.northwestern.edu/userman/index.html>, letzter Zugriff am 10.4.2014

¹⁷ Im Web unter <http://voyant-tools.org/>, letzter Zugriff am 10.4.2014

Texte dargestellt. Es besteht die Möglichkeit aus dieser Wortliste bestimmte auszuwählen und im Kontext (linke und rechte Nachbarn) zu betrachten (4) sowie eine genaue Häufigkeitsverteilung im Text (5) anzuzeigen.

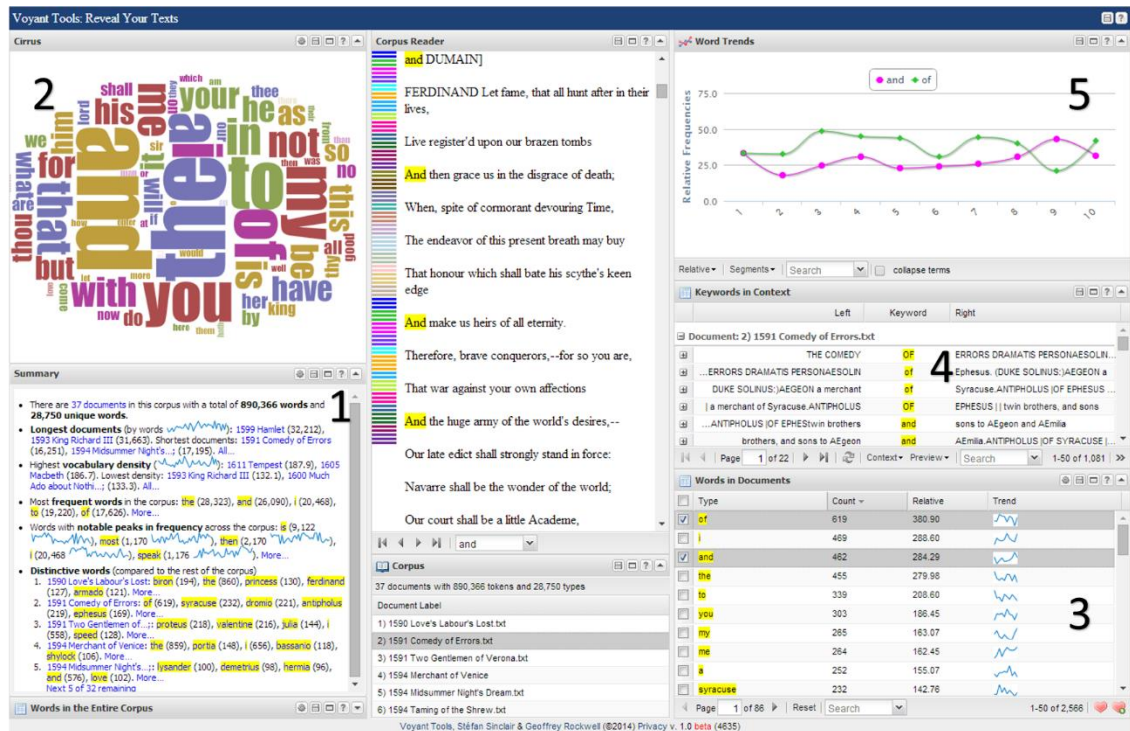


Abbildung 2: Benutzeroberfläche der Voyant Tools

Ein Zugriff auf Sinneinheiten wie Charaktere oder Zeitabschnitte wie Akte und Szenen ist in diesem Werkzeug nicht direkt möglich. Da das Werkzeug nicht speziell für annotierte Texte gebaut wurde, so wie z.B. *Mandala Browser*, können derartige Funktionen nicht ohne weiteres verwirklicht werden.

3.3 WordSeer

Muralidharan & Hearst (2013) stellen mit *WordSeer*¹⁸ ein weiteres Textanalysewerkzeug vor, das den *Voyant Tools* ähnelt. Basis sind hier jedoch *TEI*-konforme Texte. Die Funktionen gehen über die Wortebene hinaus, indem mittels automatischer linguistische Verfahren eine grammatikalische Analyse vorgenommen wird. Neben *Part of Speech Tagging* verwenden die Autoren *Natural Language Processing*, um funktionale Beziehungen zwischen einzelnen Wörtern herzustellen.

¹⁸ Im Web unter <http://wordseer.berkeley.edu/>, letzter Zugriff am 10.4.2014

Auf diesen Beziehungen ruht die Kernfunktion des Werkzeugs. Eine „grammatikalische Suche“, wie die Autoren es bezeichnen, ermöglicht es dem Benutzer nach Wörtern in einer bestimmten Funktion zu suchen. Als Beispiel führen sie die Suche Wörtern durch, die durch das Possessivpronomen *her* angezeigt werden (Ergebnis z.B. *her head*). Andere Beziehungen sind z.B. „beschrieben als“, „um zu“ oder „mit“.

Ein *Word in Context*-Modul ähnlich wie bei *Voyant* wird mitgeliefert. Allerdings setzen Muralidharan & Hearst (2013) auf eine alternative Visualisierungsform, bei der Wortnachbarn in der Konkordanz baumartig gruppiert und nach Häufigkeit skaliert sind. Weiterhin können Textabschnitte frei annotiert werden und es ist möglich, die Texte „aus der Nähe“ zu lesen. Die Möglichkeit einzelne Stellen nachzuvollziehen ist trotz Konzentration auf *Distant Reading* relevant und wurde deshalb auch in *To See or Not to See* integriert.

Sowohl *WordSeer* wie auch *Voyant* wurden am Beispiel Shakespeares eingesetzt. Das folgende Projekt nimmt die gleiche Ressource und kombiniert sie mit ähnlichen Funktionen wie hier zu einem Gesamtpaket.

3.4 Open Source Shakespeare

Wie schon Busa war Johnson (2005) an der Erstellung einer Konkordanz aller Texte eines Autors interessiert. Hierzu verarbeitete er alle in der *Moby Edition* enthaltenen Werke Shakespeares und bildete daraus eine Datenbank. Diese Edition war die erste freie digitale Version und basiert auf der *Globe Edition* von William George Clark und William Aldis Wright aus dem Jahr 1864 (Johnson, 2005).

Die Seite bietet die neben einer Konkordanz die Möglichkeit, die Werke zu lesen und zu durchsuchen. Dabei können die einzelnen Teile, seien es Charaktere oder Abschnitte als Einstiegspunkte gewählt werden. Quantitative Aspekte spielen auch eine Rolle, so können die Werke z.B. nach der Anzahl der enthaltenen Sprechakte sortiert werden. In der Tradition der *Digital Humanities* bietet der Autor die gesamte Datenbank zum Download.

3.5 The Shakespeare Quartos Archive

Für die Arbeit mit einem Autor wie Shakespeare, von dem kein direkter Urtext existiert, ist die Edition entscheidend. Die Unterschiede zwischen einzelnen Editionen zu kennen und ihre Entstehung nachvollziehen zu können, ist für Wissenschaftler entscheidend.

Dieser Problematik widmet sich das *Shakespeare Quartos Archive*¹⁹. Es geht auf ein Vorhaben der *British Library* zurück, sämtliche vor 1642 gedruckten *Quartos* digital zu sammeln und vergleichbar zu machen. *Quartos* sind die früheste gedruckte Quelle Shakespeares, auf der die heutigen Editionen basieren. Zu diesem Zweck kooperierten die *Folger Shakespeare Library*, die *Bodleian Library* in Oxford, die *National Library of Scotland* und die *Edinburgh University Library* (British Library). Beim Vorgängerprojekt *Shakespeare in quarto*²⁰ kann der Nutzer 107 verschiedene solcher *Quartos*, aufgeteilt auf 21 Stücke, nebeneinander anzeigen. Der Vergleich basiert dabei auf Fotos der jeweiligen Ausgabe. Abbildung 3 zeigt den Vergleich von *Hamlet* im ersten *Quarto* von 1603 (links) und dem zweiten *Quarto* von 1604 (rechts), die aus unterschiedlichen Quellen stammen.

Hier wird eine entscheidende Einschränkung deutlich: Der Text liegt nur als Bild vor. Für eine Weiterverarbeitung des Textes ist dieses Format ungeeignet. Es kann zwar zwischen den einzelnen Seiten geblättert werden, ein genauerer Zugriff auf kleinere Einheiten (wie Zeilen, Wörter, Charaktere, etc.) oder eine einfache Suche nach Zeichenketten ist deshalb jedoch nicht möglich. Dies verdeutlicht einerseits noch einmal den Wert der annotierten Texte, die für dieses Projekt zur Verfügung stehen. Andererseits illustriert es gleichzeitig die Verfügbarkeit von geeigneten Texten als Grundproblem der *Digital Humanities*.

¹⁹ Im Web unter <http://www.quartos.org/>, letzter Zugriff am 25.4.2014

²⁰ Im Web unter <http://www.bl.uk/treasures/shakespeare/homepage.html>, letzter Zugriff 20.4.2014

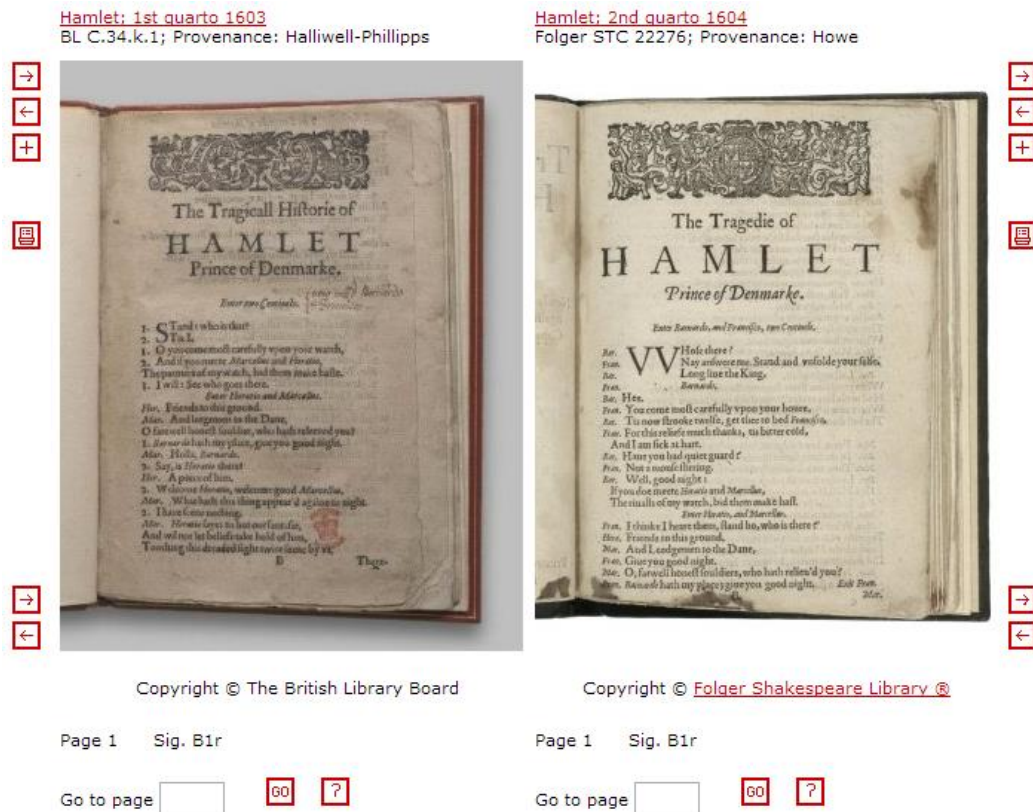


Abbildung 3: Shakespeare in quarto. Vergleich zweier Hamlet-Drucke

Die hierauf aufbauende US-UK-Kooperation²¹ *The Shakespeare Quartos Archive* will genau diese Einschränkung aufheben, indem die Texte aus den Bildern extrahiert und annotiert werden sollen. Hierzu wurden die Stücke in ein TEI-konformes Format gebracht.²² Auf Basis dieser Daten entstand ein Analysewerkzeug im Browser²³, das bisher nur die *Quartos* zu *Hamlet* umfasst. Allerdings können mehr als zwei Ausgaben gleichzeitig bearbeitet werden. Es besteht weiterhin die Möglichkeit einzelne Ausschnitte (Bilder und Text) herauszuheben und so zur weiteren Bearbeitung vorzuhalten. Eine Suchfunktion rundet die Funktionsliste ab.

Der Vergleich zweier Editionen wurde für *To See or Not to See* auch kurz angedacht aber schließlich aus Überlegungen zu Aufwand und Umfang verworfen. Dazu sollten die weiter oben vorgestellten Daten von *Open Source Shakespeare* und die von *Folger* be-

²¹ Siehe hierzu insbesondere <http://www.quartos.org/info/links.html> und <http://www.quartos.org/info/about.html>, letzter Zugriff am 18.4.2014

²² Die Annotation ist dabei nicht so detailliert wie die der Folger-Texte. Die kleinste Einheit bilden Zeilen. Die zugrundeliegenden XML-Daten sind im Werkzeug einsehbar und wurden für eine kurze Einsichtnahme heruntergeladen.

²³ Das Programm läuft unter <http://www.quartos.org/main.php>, letzter Zugriff am 18.4.2014

nutzt werden. Eine erste Analyse zeigte, dass die Daten grundsätzlich kompatibel wären, bewegte sich der Funktionsumfang im Bereich dessen, was das *Quartos Archive* anbietet.

3.6 Simulated Environment for Theatre

Die neueste Anwendung mit Verwandtschaft zum vorliegenden Projekt ist das *Simulated Environment for Theatre*²⁴ (SET) von Roberts-Smith et al. (2013). Dabei handelt es sich wie beim *Mandala Browser* um ein Projekt aus dem Kreis des *Experimental Reading Workshop* der Universität Alberta. Die Anwendung bringt eine neue Dimension in die Textvisualisierung. Die Autoren haben das Ziel mittels einer 3D-Visualisierung eines Stücks auf einer virtuellen Bühne, die sich an einer Zeitachse entlang entwickelt, die westliche Theatertradition erfahr- und erforschbar zu machen. Abbildung 4 zeigt die virtuelle Bühne auf der die aktuell anwesenden Personen mittels Sprechblasen das Stück wiedergeben.

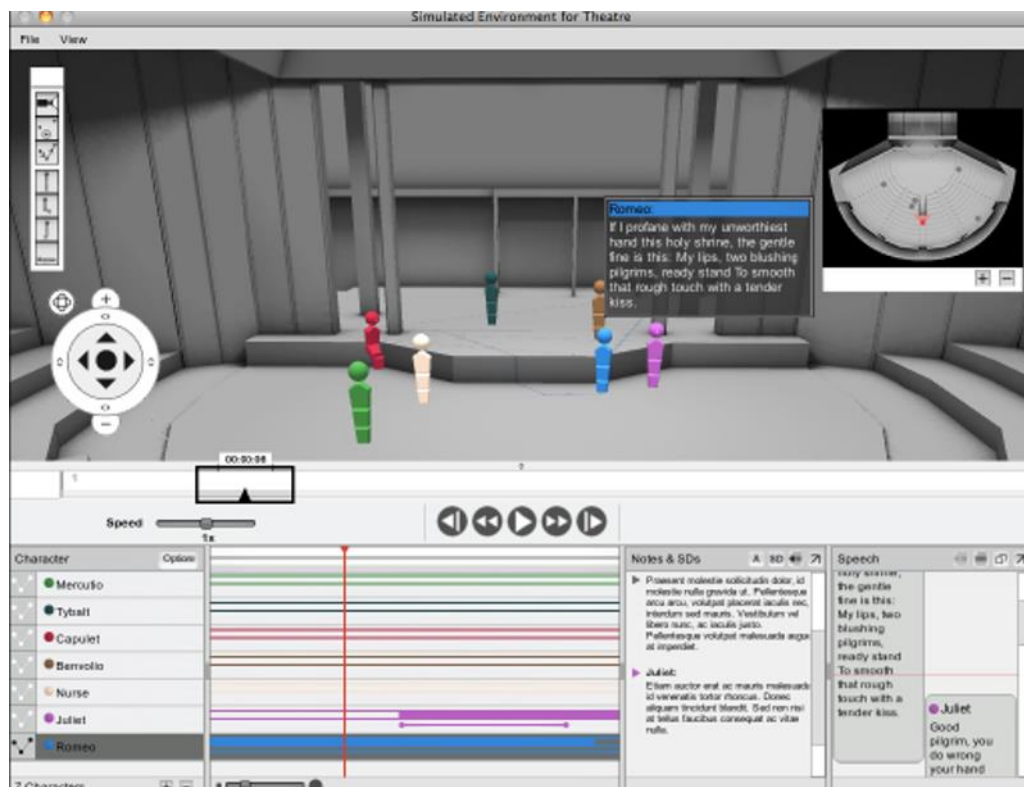


Abbildung 4: Simulated Environment for Theatre (Roberts-Smith, et al., 2013)

²⁴ Im Web unter <http://humviz.org/set/>, letzter Zugriff am 21.4.2014

SET ist weniger als automatische Visualisierung von Dramen gedacht sondern gibt dem Benutzer die Möglichkeit ein geladenes Stück (XML-Format) für die dreidimensionale Darstellung zu annotieren. Die Länge von Sprechakten, Bühnenweisungen und Bewegungen im Raum kann auf der Zeitachse frei festgelegt werden. Die Bewegungen der Figuren und ihr Aussehen werden ebenfalls vom Benutzer festgelegt. Es besteht die Möglichkeit das gesamte Stück zu editieren, indem der Text verändert oder Personen hinzugefügt bzw. entfernt werden können. Insofern eignet es sich zur Visualisierung bestehender Stücke oder zum Erstellen einer Simulation für ein neues Drama. Mithilfe der frei bewegbaren Kamera besteht die Möglichkeit das Stück aus Sicht eines Schauspielers oder Zuschauers zu verfolgen.

Das *SET*-Projekt weist einige Ähnlichkeiten zu *To See or Not to See* auf. So werden Personen farblich kodiert, auf einer Zeitachse werden die Aktivitäten für jeden Charakter angezeigt und es besteht die Möglichkeit der freien Navigation durch die Theaterstücke. Inwiefern sich *To See or Not to See* aber von den bislang vorgestellten Projekten unterscheidet, soll im nächsten Abschnitt deutlich gemacht werden.

3.7 Abgrenzung von *To See or Not to See*

Während bei *SET* die Visualisierung der Theatererfahrung in den Vordergrund gestellt wird, haben der *Mandala Browser* und *To See or Not to See* den Zweck, die Dramen analysierbar zu machen. Geschieht dies beim *Mandala Browser* auf Basis von Graphen gruppierter Elemente, wird hier eine quantitative Visualisierung von Einzelaspekten vorgenommen. Die Visualisierungen stehen im Vergleich zu allen anderen Anwendungen im stärker im Vordergrund und basieren auf komplexeren Daten. In beiden Fällen handelt es sich insofern um einen *Rich-Prospect Browser*, da alle Informationseinheiten überblicksartig dargestellt werden. Beim *Mandala Browser* sind diese Einheiten frei wählbar, bei *To See or Not to See* sind es die einzelnen Sprechakte.

Eine weitere Ähnlichkeit zum *Mandala Browser* ist die Datenbasis. In beiden Fällen werden Texte im XML-Format verwendet, wobei hier die Annotation in den Folger-Texten für die Funktionalität mit ausschlaggebend sind. Von den Textanalysewerkzeugen arbeiten nur die von *Voyant* mit nicht-annotiertem Text, was eine funktionale Limitierung mit sich bringt. Die anderen verwenden Texte auf XML- bzw. TEI-Basis, wobei insbesondere bei *WordSeer* und *SET* von einer Spezifizierung der Dokumente ausgegangen

werden muss, die nicht den Standards entspricht. Dies gilt auch für die Texte von Folger, bei welchen das standardmäßige *TEI-Tagset* geringfügig erweitert wurde. In beiden Fällen sind demnach Auswirkungen auf ihre Anwendbarkeit für beliebige Texte zu erwarten. Die Auswirkungen auf den Anspruch von *To See or Not to See*, beliebige Dramen untersuchbar zu machen, die in einem adäquaten Format vorliegen, werden später diskutiert (Abschnitte 4.8.1 und 5.3).

Voyant Tools, *Open Source Shakespeare* und das *Shakespeare Quartos Archive* laufen, so wie die hier vorgestellte Anwendung, im Webbrowser. Alle Projekte nutzen hierzu Webstandards, was aus technischer Sicht eine weitestgehend unproblematische Nutzung ermöglicht. Beim *Quartos Archive* wird eine *Java*-Applikation für den Browser verwendet, die nicht standardmäßig auf jedem System sofort unterstützt wird. *SET*, der *Mandala Browser* und *WordSeer* sind hingegen native Programme, die heruntergeladen und lokal installiert werden müssen. Dies schafft zwar eine Einstiegshürde ist jedoch durch den spezifischen Funktionsumfang begründet. Dadurch dass bei *To See or Not to See* keine serverseitige Software eingesetzt wird, kann der Benutzer sämtliche Quelldateien einsehen²⁵ – bei beinahe allen anderen Projekten ist dies nicht im vollem Umfang möglich (außer *Open Source Shakespeare*).

Eine Sonderrolle spielt das *Shakespeare Quartos Archive*. Sein Vorläufer zeigt einerseits die technischen Limitierungen auf, die fehlende Annotationen zur Folge haben. Andererseits liegt der Fokus, im Gegensatz zu den meisten anderen Projekten und zu hier, auf dem Vergleich verschiedener Versionen des gleichen Texts. Es handelt sich um eine Spezialanwendung für den Vergleich von Shakespeare-Editionen. Der *Mandala Browser*, die *Voyant Tools* und *WordSeer* sind nicht auf einen Autor oder eine Textsorte beschränkt. Sie ermöglichen einen alternativen Blick auf Literatur. *To See or Not to See* ist zwar inhaltlich ähnlich beschränkt wie *Open Source Shakespeare*, geht aber im Sinne der Textvisualisierung weiter, indem datengetriebene Visualisierungen eines Einzeltexts im Vordergrund stehen. Inwieweit hier also von *Distant Reading* gesprochen wird, soll später diskutiert werden (Abschnitt 5.2). Entstehung, Funktionsumfang und –weise von *To See or Not to See* folgen im nächsten Teil.

²⁵ Eine Ausnahme ist das serverseitige *Perl*-Skript, das die Diagramme und Graphen, die in SVG vorliegen in PNG-Bilder umwandelt (Abschnitt 4.8.3). Das Original-Skript ist jedoch quelloffen verfügbar.

4 Konzeption und Implementierung von „To See or Not to See“

To See or Not to See ist eine browserbasierte Webanwendung zur freien Navigation durch die Dramen Shakespeares mit dem Ziel der Sichtbarmachung quantitativer Aspekte. Im Zentrum der Anwendung steht die *Rich-Prospect Browsing* Idee: Alle für ein Drama wichtigen Elemente werden gleichzeitig angezeigt. Die lineare, textbasierte Darstellung wird zu diesem Zweck anhand zweier Achsen aufgebrochen: Der zeitlichen Struktur auf der x-Achse und den Personen auf der y-Achse. Der gesamte Inhalt, d.h. Text (in einzelnen Sprechakten) und Bühnenanweisungen, ist an diesen Strukturmerkmalen orientiert.

Ausgehend von der Startansicht (Abbildung 5) hat der Benutzer die Möglichkeit mit allen diesen strukturellen und inhaltlichen Elementen zu interagieren. Mit einem Klick auf eines dieser werden in den Detailfenstern dazugehörige Informationen angezeigt, wie es für *Rich-Prospect Interfaces* typisch ist. Das Fenster *speech* zeigt dabei den gesprochenen Text für das ausgewählte Element an. Im Fenster *metrics* werden kontextspezifische quantitative Daten und andere, weiterführende Informationen angezeigt.

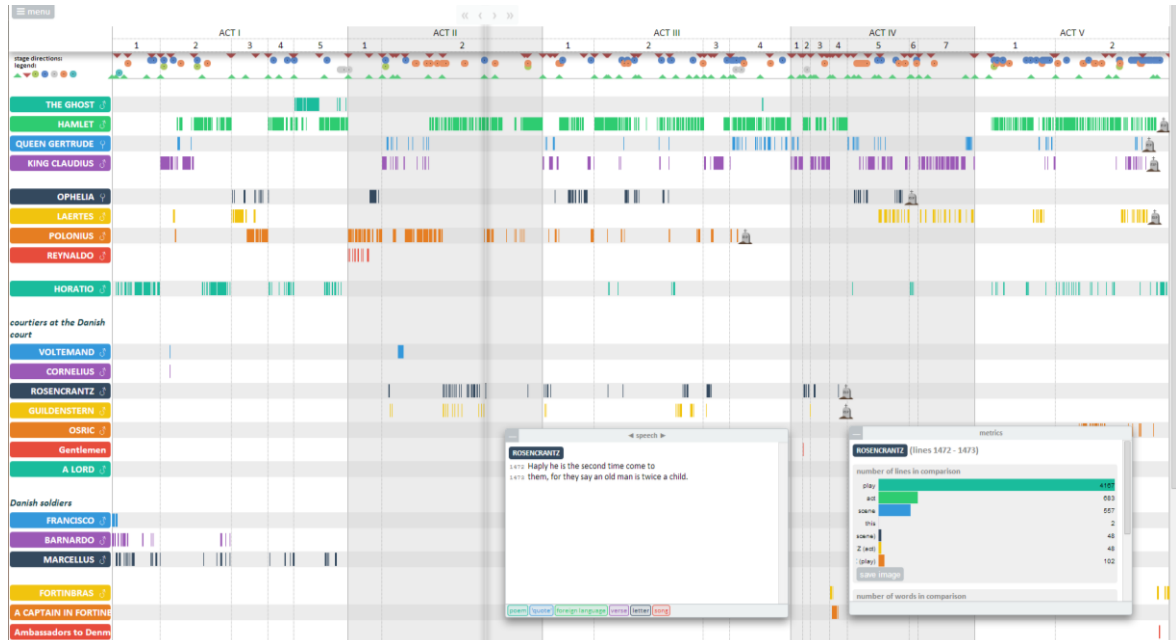


Abbildung 5: Überblicksansicht und Detailfenster (unten mittig, unten rechts)

4.1 Zeitachse

Die zeitliche Strukturierung wird auf Basis der Zeilenanzahl der einzelnen Abschnitte (Akte und Szenen) vorgenommen. Somit dienen die einzelnen Zeilen des Werks als Abstraktion der Zeit, die bei der Aufführung auf der Bühne vergeht. Die Achse verläuft horizontal von links nach rechts. Im Vergleich zur Erzählzeit bietet dieser Ansatz zwei wesentliche Vorteile. Zum einen soll das Werkzeug einen objektiven Blick auf den Text an sich ermöglichen und vom Inhalt abstrahieren. Zum zweiten ergibt sich die ganz praktische Einschränkung, dass mit den vorliegenden Grunddaten und heutigen Mitteln eine Ermittlung der erzählten Zeit unrealistisch ist.

Entlang der Zeitachse kann ein Schieberegler (in Abbildung 5 auf Höhe von Akt 2, Szene 2; Detailansicht in Abbildung 16) dazu genutzt werden, durch die Sprechakte des Stücks zu navigieren. Er zeigt die aktuelle Stelle im Stück an, sobald ein Sprechakt ausgewählt worden ist. Es besteht die Möglichkeit mithilfe des Schiebereglers einen bzw. zehn Sprechakte (einfache bzw. doppelte Pfeile) vor oder zurück zu gehen. Der Regler kann außerdem mit der Tastatur bedient werden. Den einzelnen Knöpfen sind die Tasten „1“ (10 nach links), „2“ (1 nach links), „3“ (1 nach rechts) und „4“ (zehn nach rechts) zugeordnet.

4.2 Personenachse

Als zweites Strukturierungsmerkmal werden die Personen des Stücks verwendet. Sie sind auf der linken Seite aufgelistet. Die Liste richtet sich strikt nach denen im Ausgangstext aufgeführten Charakteren. Diese werden nach der im Ausgangsformat ebenfalls vorgegebenen Zuordnung zu Gruppen, Familien o.ä. zusammengefasst. So zum Beispiel *Francisco*, *Barnardo* und *Marcellus*, die dänischen Soldaten bei *Hamlet* (Abbildung 5). Mehrere Schauspieler können im *TEI*-Format dabei unter einem einzelnen Charakter aufgeführt (z.B. *Two Messengers* und *Sailors* in *Hamlet*) werden.

Alle Einzelpersonen sind dabei über einen eindeutigen Schlüssel identifiziert. Diese Schlüssel werden in der weiteren Modellierung dazu eingesetzt, in Sprechakten, Bühnenanweisungen, Akten und Szenen auf Personen zu verweisen. Für eine automatisierte Datengewinnung, wie sie hier vorgenommen wird, ist diese Eigenschaft der Folger-Texte wichtig. Am Beispiel von Bühnenanweisungen wird dies weiter unten exemplarisch im Detail erklärt.

4.3 Inhalt

Die beiden Achsen spannen den zweidimensionalen Raum auf, in dem der Text, aufgebrochen in einzelne Sprechakte, platziert wird. Diese werden nicht direkt dargestellt, sondern auf ihre relative Länge in Zeilen reduziert. Diese Abstraktion ist im Sinne des *Rich-Prospect Browsing*. Die Breite der einzelnen Sprechakte richtet sich also nach der gleichen Größe wie die strukturellen Elemente (Akte und Szenen). Je länger ein Charakter am Stück spricht, desto breiter ist der farbige Streifen an der entsprechenden Stelle im Text. Die Farbkodierung und vertikale Orientierung helfen bei der Identifizierung der jeweiligen Sprecher. Die minimale Breite von nur eine Zeile langen Sprechakten liegt bei einem Pixel – ansonsten wären diese gar nicht sichtbar. Da sich die Breite der Anwendung nach der Browserfenstergröße richtet, ist die Darstellungsgenauigkeit durch die Auflösung der Anzeige begrenzt. Bei einer Bildschirmauflösung von 1920 Pixeln in der Horizontalen und einem Stück wie *Hamlet* mit 4167 Zeilen wären die kleinsten Einheiten etwas weniger als $\frac{1}{2}$ Pixel breit, zieht man Ränder und die Charakterleiste ab. Insofern ist die künstlich festgelegte Mindestbreite sinnvoll, um weiterhin alle Elemente anzeigen zu können.

Durch die Mindestbreite der Sprechakte kann teilweise der Eindruck zusammenhängender, langer Sprechhandlungen entstehen oder sich diese im Extremfall überlappen. Dies verfälscht einerseits die Darstellung wird andererseits aber durch die Interaktionsmöglichkeiten wieder relativiert. Da der Benutzer die Möglichkeit hat mit jedem Element auf unterschiedliche Arten einzeln zu interagieren, ist schon nach kurzer Benutzungsdauer ein Bewusstsein für diese Darstellungsproblematik vorhanden. Für das Gelingen der Interaktion, also die Auswertung von spezifischen Aspekten der Stücke, ist dies somit unerheblich. Für die Zukunft ist eine Möglichkeit zur Vergrößerung bzw. Verkleinerung der Ansicht geplant, die möglichst unabhängig von der Auflösung des Benutzers funktioniert und die gesamte Kollektion zeigen kann.

4.4 Bühnenanweisungen

Alle Bühnenanweisungen sind ebenfalls entlang der Zeitachse orientiert und nicht direkt den betreffenden Personen zugeordnet. Diese Vorgehensweise wurde aus zwei Gründen gewählt. Einerseits sind nicht alle Bühnenanweisungen direkt auf eine oder

mehrere Personen bezogen. Würden also alle personenbezogenen Anweisungen den betroffenen Charakteren visuell zugeordnet, blieben einige Anweisungen übrig. Diese müssten ohnehin ohne Personenzuordnung angezeigt werden, weshalb ein eigener Ort zur Anzeige in jedem Fall notwendig ist. Die daraus folgende Zersplitterung aller Bühnenanweisung auf verschiedene Bereiche würde nicht zur Übersichtlichkeit beizutragen.

Andererseits ist es sinnvoll die Anweisungen von Personen losgelöst zu zeigen, da diese nur auf Personen bezogen sein können, aber nie von Charakteren „angestoßen“ werden. Sie werden den Personen von extern auferlegt, sie gehören in den Bereich Inszenierung. Insofern sind sie nicht Teil des Texts und werden auch nicht in der Textkollektion angezeigt.

Welche Personen von einer Bühnenanweisung betroffen sind ist grundlegend im Markup der Ausgangstexte hinterlegt. Sie werden jedoch nur aufgelistet. Welcher konkrete Teil der Anweisung die einzelnen Personen betrifft, ist demnach nicht automatisch nachvollziehbar. Dennoch konnte diese Auszeichnung für das Werkzeug gewinnbringend eingesetzt werden, wie das Beispiel in Abschnitt 4.6.4 zeigt.

Auf Basis dieser Annotationen ist es prinzipiell möglich, für jeden beliebigen Zeitpunkt im Stück automatisch anzugeben, wer auf der Bühne präsent ist. Bisher ist die kleinste Einheit, für die die Anwesenheit von Personen anhand von Sprechakten in *To See or Not to See* ermittelt wurde, die Szene. Verlässt z.B. eine Person die Bühne während einer Szene, wird dies in der späteren Auswertung für diese Szene bzw. den umgebenden Akt nicht berücksichtigt. Eine genauere Strukturierung nach diesen „französischen“ Szenen (Roberts-Smith, et al., 2013), kann für die Zukunft eine interessante Erweiterung darstellen.

4.5 Visualisierung quantitativer Daten

Die wichtigste Komponente von *To See or Not to See* sind die Visualisierungen quantitativer Daten zu den einzelnen Aspekten der Theaterstücke. Für Charaktere, Akte und Szenen sowie alle Sprechakte gibt es verschiedene Grafiken, die unterschiedliche quantitative Eigenschaften hervorheben.

Es kommen zweierlei Diagramme zum Einsatz: Erstens Balkendiagramme, welche Häufigkeiten von Zeilen, Wörtern, und Bühnenanweisungen in Relation zum jeweiligen

Kontext setzen. Zweitens werden Tortendiagramme (als Ringe dargestellt) verwendet, um bei den Abschnitten Sprecheranteile zu visualisieren. Die Balkendiagramme werden bei der Untersuchung von Abschnitten, Sprechakten und Charakteren eingesetzt. Bei Akten und Szenen wird deren Länge, in Wörtern und Zeilen, mit der Länge des Stücks bzw. des übergeordneten Abschnitts vergleichbar gemacht. Bei Personen zeigen sie die relative Verteilung der Sprechaktivität auf die einzelnen Teile des Stücks. So ist z.B. für jeden Charakter angegeben, wie viele Zeilen (bzw. Wörter) dieser in seinen aktiven Szenen spricht. Bei den Sprechakten kommen zwei Balkendiagramme zum Einsatz, die die Länge des ausgewählten Dialogabschnitts in Relation zum umgebenden Kontext (Szene, Akt und Stück) setzen. Gleichzeitig wird Bezug auf die sprechende Person genommen und der Anteil angezeigt, den dieser Sprechakt für diese Person im Stück ausmacht.

Das Verhältnis von Sprechanteilen in einem Akt oder einer Szene wird mithilfe der Ringdiagramme dargestellt. Diese zeigen die Anzahl der gesprochenen Zeilen bzw. Wörter für die aktiven Personen. Dieses Gewicht einzelner Charaktere in einem Abschnitt ist gleichzeitig in anderer Form in den Personennetzen enthalten (Größe der Knoten).

Die Netze kommen in den Auswertungen für Charaktere, Akte und Szenen zum Einsatz. Die Graphen zeigen Verbindungen zwischen einzelnen Personen und deren Gewicht (gemessen an den Zeilen) im Kontext an. Die Verbindungen werden auf Basis von in gemeinsamen Szenen beteiligten Charaktere hergestellt. Einen ähnlichen, jedoch manuellen Ansatz verfolgt Moretti (2013, S. 214), wobei er sich auf Vorarbeiten²⁶ anderer Autoren mit dem gleichen Ansatz stützt. Das Personennetz einer Einzelperson zeigt alle mit ihr über die Dauer des ganzen Stücks verbundenen Charaktere. Die Stärke ihrer Verbindung wird durch die Nähe der Knoten und die Dicke der Kanten dargestellt. Die Größe eines Knoten zeigt an, welchen Sprechanteil die Person relativ zu den anderen im gegebenen Kontext besitzt. Die gleichen Eigenschaften gelten für die Personennetze, welche insbesondere bei Akten aber auch bei Szenen wichtige Informationen übermitteln. Diese Netze zeigen nicht nur die Verbindungen einer Person mit ihrem Umfeld, sondern alle Verbindungen aller in diesem Abschnitt aktiven Charaktere. Es entstehen

²⁶ Ob zwei Personen, die gleichzeitig auf der Bühne sind tatsächlich miteinander sprechen ist dort unerheblich. Gleichzeitige Anwesenheit auf der Bühne genügt. Automatisiert zu erheben, welche Charaktere in Sprechakten adressiert werden, könnte in Zukunft unter anderem durch das Heranziehen von Bühnenweisungen vom Typ *delivery* versucht werden, ist aber in diesem Fall eher von der Inszenierung abhängig als von den eigentlichen Texten.

somit die gleichen Netze, wie sie Moretti in *Network Theory, Plot Analysis in Distant Reading* (2013) am Beispiel *Hamlets* manuell erstellt. Der Unterschied liegt darin, dass die Netze automatisch auf Basis des annotierten Textes gebildet werden. Moretti weist nach, welche interpretative Kraft in derartigen Netzen steckt. Aus diesem Grund wurden sie in das Werkzeug integriert.

Für ihre Darstellung kommt die *Force Graph*-Komponente von *d3.js* zum Einsatz. Dabei handelt es sich, wie bei den oben genannten Diagrammen, um eine vordefinierte Visualisierungsmethode der Bibliothek. Um sie einzusetzen, ist es Aufgabe des Programmierers einen Adapter erstellen, der die Daten und ihre Verwendung durch die Bibliothek festlegt. Werden diesem Modul schließlich adäquat modellierte Datenobjekte übergeben, wird ein Graph erstellt, dessen Form durch die Berechnung abstrakter Kräfte bestimmt wird. So können sich z.B. einzelne Knoten auf Basis ihrer Eigenschaften entweder abstoßen oder anziehen. Die Berechnung der Kantenlänge beruht auf demselben Prinzip: Der Programmierer legt fest, welche Eigenschaften der Knoten für die Darstellung des Graphen relevant sind und wie sie sich auf diese auswirken. Bei *To See or Not to See* ist dies die Anzahl der den Personen gemeinsamen Szenen. Auf das tatsächliche Aussehen der Graphen hat man somit nur indirekt Einfluss. Jedes neu generierte Netz weist optische Unterschiede auf, obwohl es auf Basis derselben Daten berechnet worden ist. Um eventuellen Verzerrungen, Unlesbarkeit oder anderen suboptimalen Ergebnissen entgegen zu wirken, können die einzelnen Knoten mit der Maus verschoben werden – das Netz wird durch Funktionen von *d3.js* automatisch an die neue Position angepasst. An dieser Stelle wurde in dieses Standardverhalten eingegriffen und festgelegt, dass Knoten, sobald sie vom Benutzer bewegt worden sind, auf der Leinwand haften bleiben. Somit ist es möglich den Graphen nach Belieben zu manipulieren, um bestimmte Konstellationen sichtbar zu machen oder Eigenschaften hervorzuheben. Die Kraft, die hinter der berechneten Kantenlänge steckt, wird dabei ausgehebelt – Personen können näher zusammengedrückt werden, um sie z.B. zu Gruppen zusammenzufassen. Diese Erweiterung wurde vorgenommen, um die Objektivität der Darstellung nicht dem Algorithmus zu überantworten, sondern dem Benutzer.

4.6 Anwendungsszenarien und Funktionen

Die zentralen Komponenten (Übersichtsansicht, Detailfenster, Visualisierungen und Interaktionsmöglichkeiten) bieten dem Nutzer verschiedenste Möglichkeiten, Informationen über die verfügbaren Stücke zu erfahren. Im Folgenden werden die einzelnen Funktionen anhand isolierter Nutzungsszenarien gezeigt.

4.6.1 Informationen über Charaktere abrufen

Will ein Benutzer mehr über einen einzelnen Charakter eines Stücks erfahren, so kann er dies mittels eines Klicks auf dessen Namen in der am linken Bildschirmrand befindlichen Liste anstoßen. Es werden sodann im Detailfenster *speech* sämtliche Sprechakte dieser Person bzw. Gruppe angezeigt. Weiterführende Informationen erscheinen im Fenster *metrics*, wie Abbildung 6 am Beispiel von *Ophelia* in *Hamlet* zeigt:

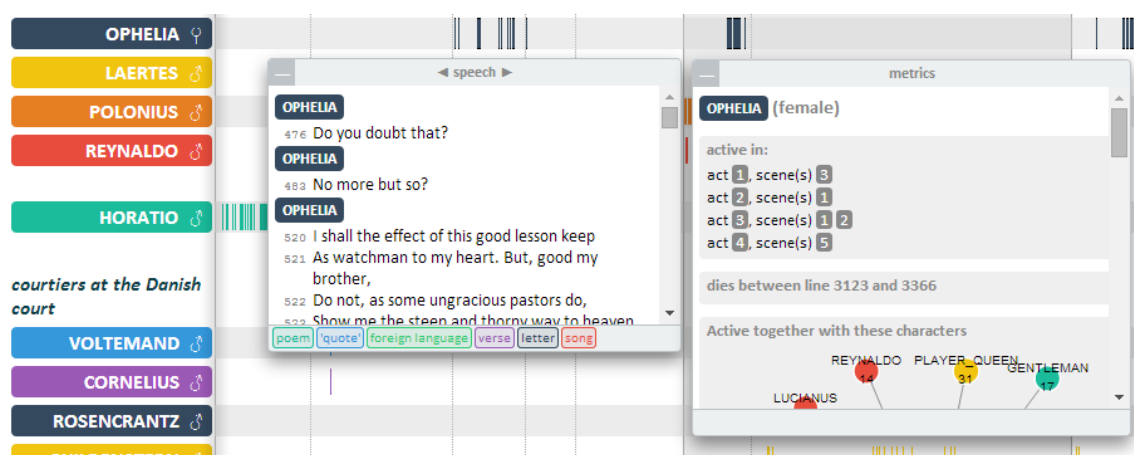


Abbildung 6: Detailinformationen zum Charakter *Ophelia* in *Hamlet*

Im Textfenster sind die einzelnen Sprechakte durch das wiederholte Auftreten des Namens inkl. farblicher Markierung voneinander getrennt. Links neben jeder Zeile wird ihre jeweilige laufende Nummer angezeigt. Im Beispiel ist zu sehen, dass *Ophelia* in diesem Teil in einem Dialog steht, da ihre Sprechakte nur kurz unterbrochen werden. Die Übersichtsansicht (im Bild verdeckt) deutet an, dass sie sich im Gespräch mit *Laertes* und *Polonius* befindet.

Am unteren Rand des Fensters ist eine Legende abgetragen, die anzeigt mit welchen farbigen Umrandungen bestimmte Teile des Texts markiert sind. Dazu gehören Textpassagen, die Gedichte (*poem*), Zitate (*quote*), fremdsprachige Ausdrücke (*foreign language*), Verse (*verse*), Auszüge aus Briefen (*letter*) und Gesang (*song*) enthalten und von Folger in den Ausgangstexten entsprechend annotiert worden sind. In Zeile 3087

singt *Ophelia* „They bore him barefaced on the bier“ und weist ihren Gegenüber in den Zeilen 3093 bis 3096 beim Singen an, indem sie den Text zitiert (blau markiert, einfache Anführungszeichen zur Verdeutlichung hinzugefügt), wie in Abbildung 7 zu sehen ist.

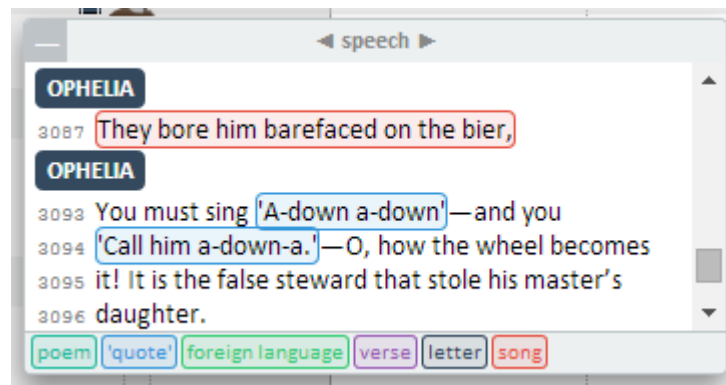


Abbildung 7: Gesang und Zitat in Sprechakten *Ophelias*

Neben der Funktion, den gesamten Text einer Person anzuzeigen, können so Charaktere identifiziert werden, die z.B. häufig Fremdsprachen benutzen. Wann und mit wem ein Charakter im Stück aktiv ist und wie sich seine Aktivität quantitativ zum Rest des Stücks verhält, zeigt das Fenster *metrics*. *Ophelia* ist in den ersten vier Akten in den Szenen 3, 1, 1 und 2 sowie 5 am Geschehen auf der Bühne beteiligt, bevor sie zwischen Zeile 3123 und 3366 ihre Sterbeszene spielen darf (Abbildung 8).



Abbildung 8: weiterführende Informationen und quantitative Daten zum Charakter *Ophelia*

Mit einem Klick auf die Nummern der Akte und Szenen im Block *active in*, in denen ein Charakter aktiv ist, werden die weitergehenden Informationen zu dem jeweiligen Abschnitt angezeigt (s.u.). Der Tod eines Charakters ist in der Annotation entweder vage

(wie bei *Ophelia* zwischen bestimmten Zeilen) oder für eine spezifische Zeile angegeben. Dies ist davon abhängig, ob der genaue Zeitpunkt unklar ist oder von der Inszenierung abhängt (zum Vergleich: *Polonius* letzte Worte sind „O, I am slain!“ in Zeile 2466). Ist Letzteres der Fall kann auf die Zeilenangabe im *metrics*-Fenster geklickt und so zu der entsprechenden Stelle im Stück navigiert werden. In der Übersicht wird der Tod an der frühestmöglichen Stelle durch ein Grab angezeigt.

Des Weiteren wird das für den gewählten Charakter spezifische Personennetz angezeigt. Dies ist ein Graph, bei dem der Charakter mit all jenen anderen Personen verbunden ist, die mindestens in einer Szene gemeinsam mit ihm bzw. ihr auf der Bühne sind. Abbildung 9 zeigt dies am Beispiel *Ophelia*.

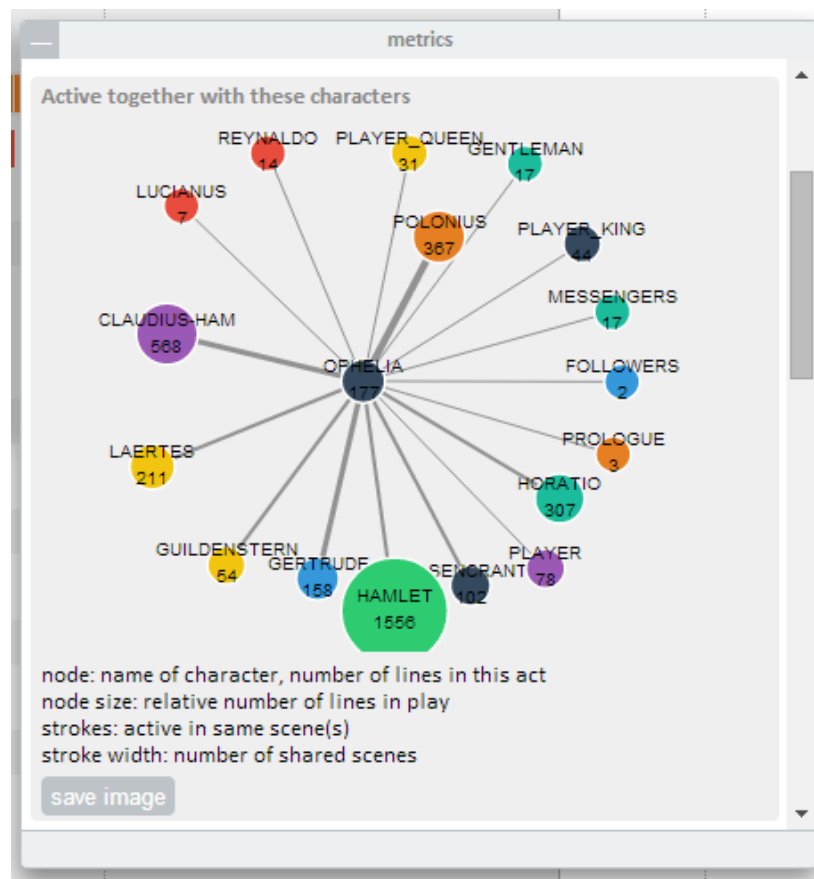


Abbildung 9: Personennetz des Charakters *Ophelia*

Die Abbildung macht deutlich, dass *Ophelia* jeweils in einer Szene mit einer ganzen Gruppe von (für sie weniger wichtigen) Personen zu sehen ist (Nebenrollen wie *Followers* und *Messengers*). Darunter *Reynaldo*, der im gesamten Stück 14 Zeilen Text hat. Wichtiger für *Ophelia* scheinen jedoch *Polonius* und *Claudius* zu sein, da sie ihr in der

Darstellung am nächsten sind und ihre Verbindungslinien (Kanten des Graphs) am breitesten sind. Die Kantenbreite ist eine Abstraktion der Häufigkeit gemeinsamer Szenen²⁷, die zusätzlich durch die Nähe der Knoten angezeigt wird. Aus demselben Grund wurden eingangs die am weitesten entfernten Personen für diesen Charakter als weniger wichtig interpretiert, da sie weiter entfernt, ihre Verbindungen weniger stark ausgeprägt sind. Die Größe der Knoten zeigt die relative Anzahl ihrer Zeilen im gesamten Stück an, die unter dem jeweiligen Knotennamen angegeben ist. Das Maximum bildet hier *Hamlet* mit 1556 Zeilen, während die *Followers* das Minimum darstellen (Die *Messengers* sind entsprechend nur geringfügig größer). Diese Darstellung hat den Vorteil, dass das relative Gewicht der gerade inspizierten Person in ihrem spezifischen Umfeld sichtbar wird.

Für die Berechnung der Graphen wurden als Grundeinheit die Zeilen des Texts für gewählt, da diese auch die Einheit sind, an der sich die Zeitachse des Stücks orientiert (siehe Abschnitt 4.1). Will ein Nutzer mehr Daten zu den durch einen Charakter gesprochenen Zeilen, so werden ihm diese direkt unterhalb des Netzes angezeigt (Abbildung 10):

²⁷ Die Anzahl gemeinsamer Szenen wird aktuell noch nicht angezeigt, ist aber in der Kante hinterlegt und kann so im Quellcode nachgelesen bzw. in Zukunft einfach ergänzt werden. Ein Pixelbreite entspricht einer gemeinsamen Szene.

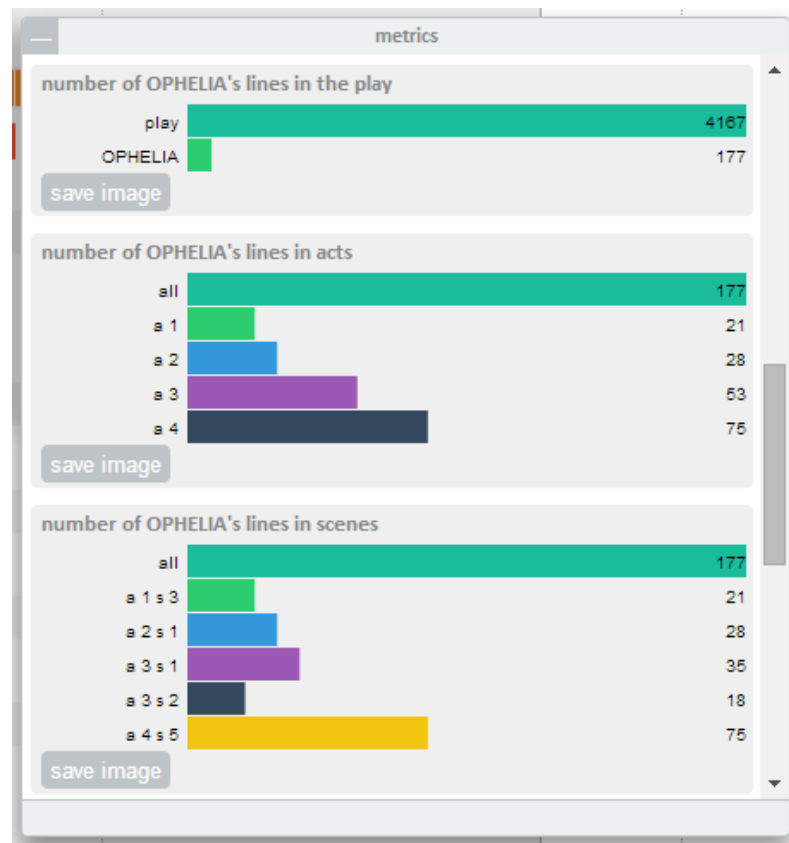


Abbildung 10: Diagramme zu den gesprochenen Zeilen des Charakters *Ophelia*

Die drei Balkendiagramme in Abbildung 10 setzen *Ophelias* Sprechanteil, gemessen in Zeilen, in Relation zum Stück (oben), zu den Akten, in denen sie aktiv ist (mittig) und ihren Einzelszenen (unten). Als Relation ist jeweils der größere Kontext gegeben (Balken über die maximale Breite). Im ersten Diagramm ist dies das Stück, das 4167 Zeilen umfasst. Im zweiten und dritten die gesammelten Zeilen *Ophelias*, die anschließend auf die Abschnitte (Akte bzw. Szenen) heruntergebrochen werden, um einen Vergleich zu ermöglichen. Im Beispiel sieht man, dass sie in Akt 4 den Großteil ihrer Dialogaktivität hat – kurz bevor sie stirbt.

Neben Zeilen ist die Anzahl der gesprochenen Wörter von Interesse. Deshalb werden darunter dieselben Diagramme mit Daten zu den Wörtern eines Charakters angezeigt (Abbildung 11):

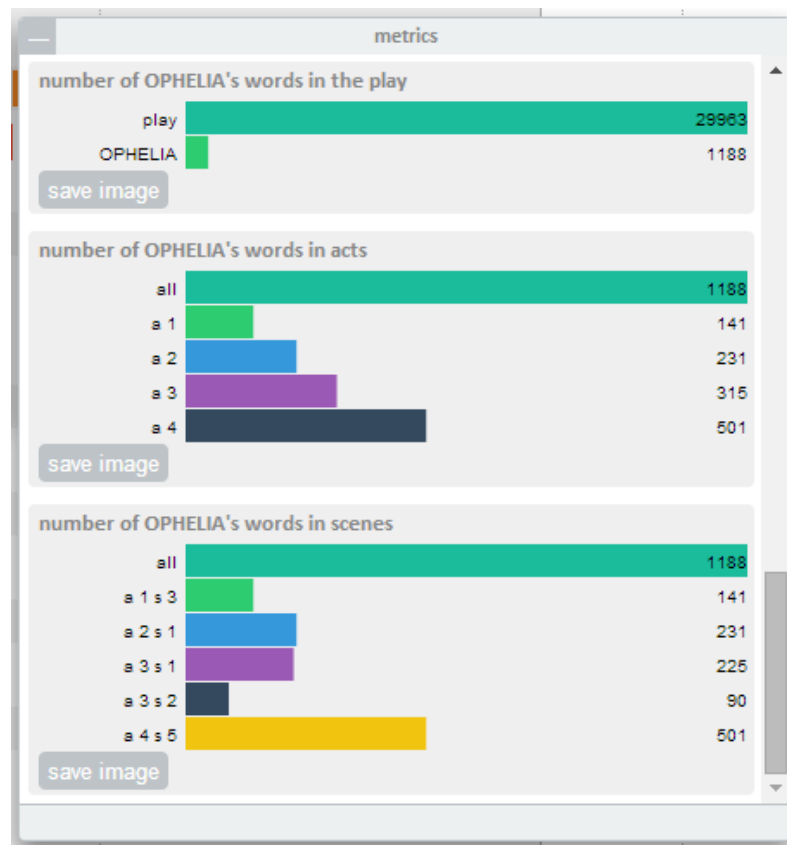


Abbildung 11: Diagramme zu den gesprochenen Wörtern des Charakters *Ophelia*

Zwar bleibt im Beispiel der Trend erhalten, dass sie von Akt 1 bis 4 immer mehr Anteile hat, jedoch besteht ein Unterschied bei Akt 2, Szene 1 und Akt 3, Szene 1. Ein Blick auf die Übersichtsansicht zeigt: In der ersten Szene des zweiten Akts hat sie längere, zusammenhängende Sprechakte. In der ersten Szene des dritten Aktes hingegen meist nur kurze, z.T. einzeilige Auftritte. Klickt man diese an, wird deutlich, dass diese häufig nur kurze Antworten im Dialog sind. Die unterschiedlichen Diagramme sind hierdurch erklärt.

4.6.2 Informationen über Akte und Szenen abrufen

Mithilfe von *To See or Not to See* wird der Zugriff auf Zeitabschnitte eines Theaterstücks möglich. Wird die Maus über den Titel des Akts bzw. der Szene bewegt, werden die inaktiven Personen in der Charakterliste am linken Rand ausgeblendet. Durch Klick auf einen Akt bzw. eine Szene in der Zeitleiste am oberen Rand werden Text und Daten zu eben jenem Abschnitt angezeigt. Analog zu den Charakteren wird hier entsprechend der Text der jeweiligen Sprecher im Akt bzw. in der Szene linear angezeigt (Abbildung 12):

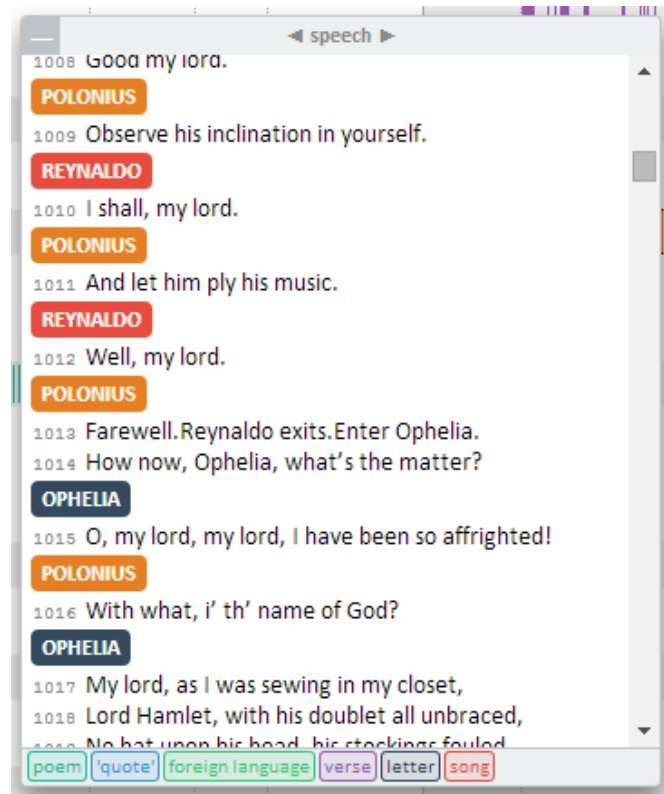


Abbildung 12: Ausschnitt aller Sprechakte in Akt 2 in *Hamlet*

Im *metrics*-Fenster werden nun kontextspezifische Informationen angezeigt. Darunter wiederum ein Personengraph, für den die gleichen Bedingungen gelten wie bei den Details zu Charakteren. Die Anzahl der Zeilen im aktuell gewählten Abschnitt wird angegeben und bestimmt die relative Größe der Knoten und gemeinsame Szenen der Charaktere sind durch die Kantenlänge und -dicke dargestellt. Inspiziert ein Nutzer eine Szene, so sind alle Kanten gleich, da die Anzeige auf eine Szene beschränkt ist. Im Beispiel (Abbildung 13) des 2. Akts in *Hamlet* gibt es zwei Hauptakteure, *Hamlet* und *Polonius*). Letzterer hat als einziger Verbindungen zu *Reynaldo* und *Ophelia*.

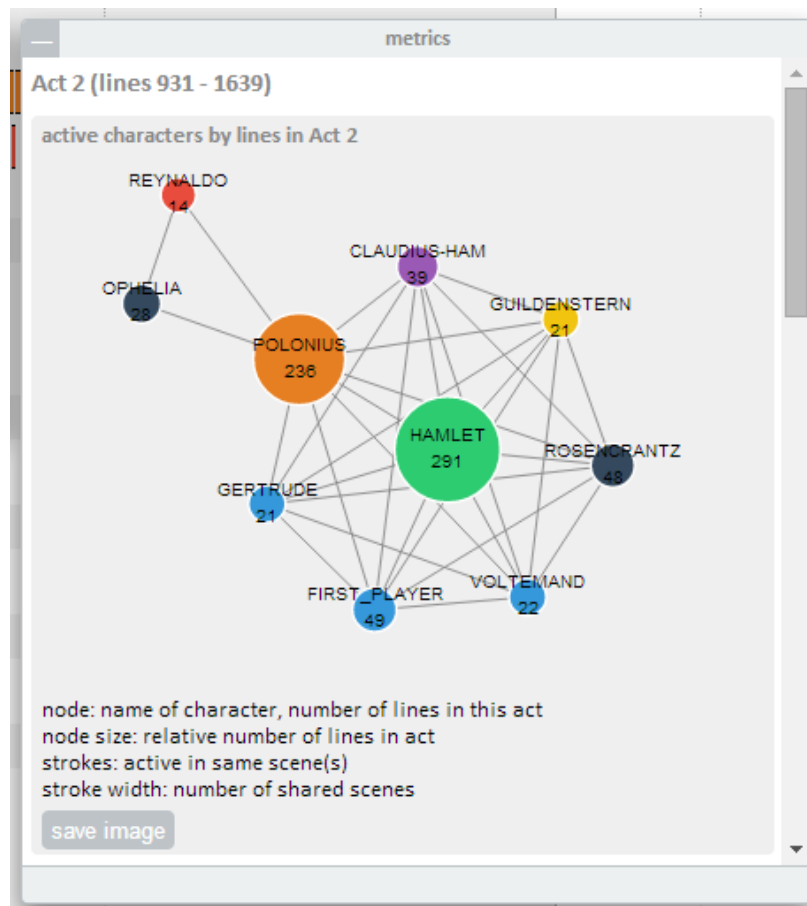


Abbildung 13: Personennetz für Akt 2 in *Hamlet*

In diesem einfachen Beispiel zeigt das Personennetz, dass es im Akt zwei Szenen gibt, die durch den Charakter *Polonius* zusammenhängen. Das automatisch gewichtete Netz wurde hier vom Benutzer manipuliert, um die Personengruppen visuell voneinander abzugrenzen, die sich um *Polonius* bilden. Hierzu wurden die Charaktere der ersten Szene enger zusammengeführt. Dies wird möglich, indem der Nutzer einen Knoten bei gedrückter linker Maustaste verschiebt und dieser anschließend nicht mehr vom automatischen Gewichtungsalgorithmus verschoben werden kann.

Diagramme zu Häufigkeiten von Zeilen, Wörtern und Bühnenanweisungen des Akts bzw. der Szene folgen unterhalb des Personengraphen. Sie zeigen diese Häufigkeiten im Kontext. Für Akte ist dies das Stück und für Szenen der übergeordnete Akt und das Stück. Am Beispiel der ersten Szene des zweiten Akts sind diese Diagramme in Abbildung 14 dargestellt.

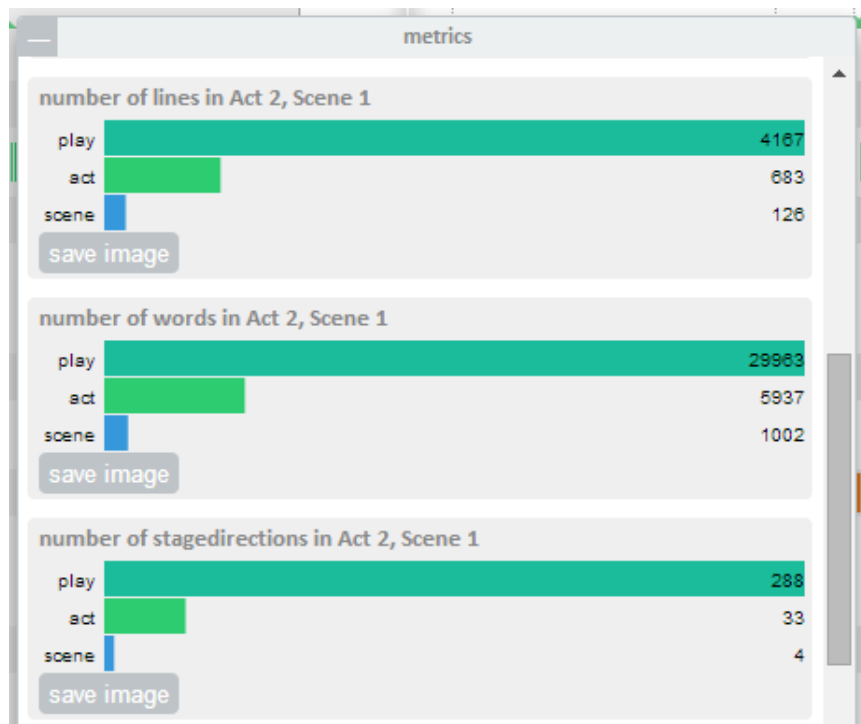


Abbildung 14: Diagramme zur Anzahl der Zeilen, Wörter und Bühnenanweisungen in Akt2, Szene 1 von *Hamlet*

Im Verhältnis zu anderen Teilen werden im gewählten Abschnitt mehr Wörter gesprochen, wie an den etwas längeren Balken im zweiten Diagramm im Vergleich zum darüberstehenden sichtbar wird. Wie weiter oben am Beispiel von *Ophelia* gezeigt wurde, kann dies z.B. darüber Auskunft geben, ob hier längere Monologe vorherrschen.

Wie sich diese Zeilen und Wörter auf die einzelnen Akteure im Kontext aufteilen, zeigt das darunter folgende Ringdiagramm. Für Bühnenanweisungen ist diese Darstellung wenig gewinnbringend, da nicht jede dieser Anweisungen auf bestimmte Personen bezogen ist.

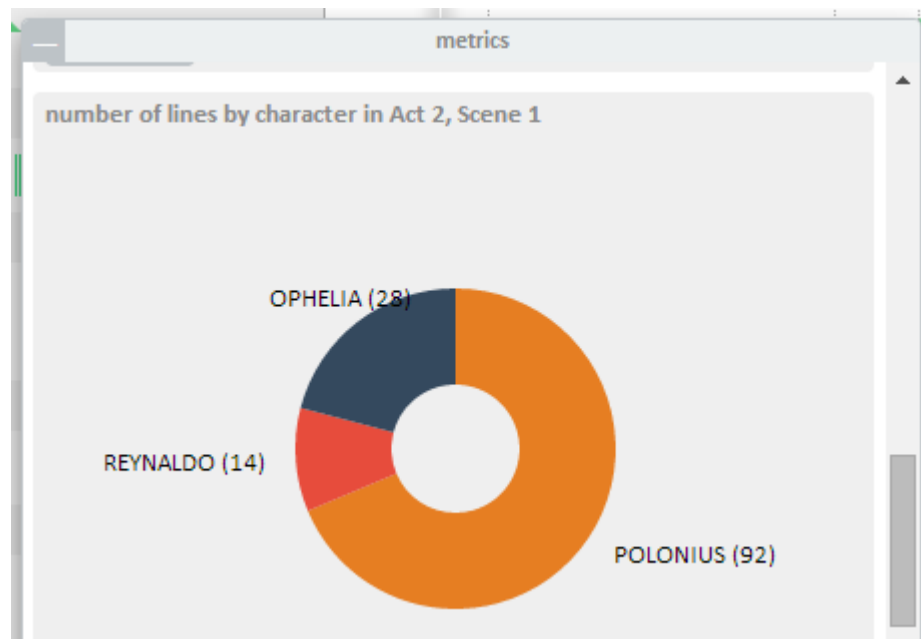


Abbildung 15: Verteilung der Zeilen auf die aktiven Personen in Akt 2, Szene 1 in *Hamlet* (analog für Wörter)

Das Beispiel in Abbildung 15 zeigt den relativen Sprechanteil der drei aktiven Charaktere in Form eines Ringdiagramms und gibt die Anzahl der gesprochenen Zeilen an. Es handelt sich hierbei um eine alternative Darstellung der Sprechanteile, die im Personennetz bereits integriert ist. Im Kontext eines Akts werden Szenengrenzen sowohl bei der Bestimmung der Knotengröße im Graphen wie auch bei der Berechnung der Ringausschnitte ignoriert.

4.6.3 Informationen zu Sprechakten abrufen

Die direkte Suche nach Zitaten ist in *To See or Not to See* noch nicht integriert. Zwar steht eine Textsuchfunktion nach Auffassung des Autors nicht zwangsweise im Widerspruch mit den Prinzipien des *Rich-Prospect Browsing*, wurde deswegen dennoch nicht priorisiert. Das wohl berühmteste – und für diese Arbeit namensgebende – Zitat aus *Hamlet* steht in Zeile 1762 und wird durch Klick auf den entsprechenden Sprechakt sichtbar. Weitere Möglichkeiten zu diesem Sprechakt zu gelangen sind einerseits die Benutzung des Schiebereglers am oberen Bildrand, der die aktuelle Position im Stück anzeigt oder andererseits der Navigationspfeile im *speech*-Fenster. Abbildung 16 illustriert den Zusammenhang.

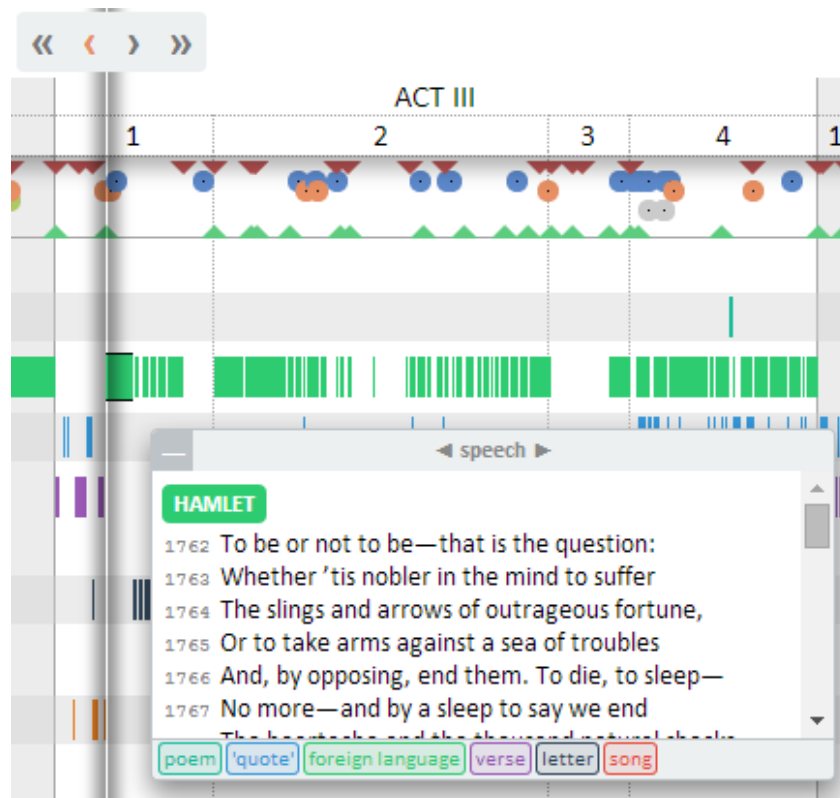


Abbildung 16: Anzeige eines einzelnen Sprechakts am Beispiel von *Hamlets* „To be or not to be“

Im Beispiel (Abbildung 16) wird der aktuelle Sprechakt durch schwarze Ränder oben und unten markiert (dieselbe Markierung geschieht bei Inspektion von Akten und Szenen) und der Schieberegler zeigt an, wo im Stück der Monolog beginnt. In die Titelleiste des Fensters sind zwei Pfeile zur Navigation nach links (im Text weiter oben) und rechts (weiter unten) integriert. Hierbei wird der vorherige bzw. nächste Sprechakt selektiert, der Schieberegler wird entsprechend verschoben. Hiermit besteht neben dem Schieberegler eine zweite Möglichkeit durch das Theaterstück zu navigieren.

Zur gleichen Zeit wird im *metrics*-Fenster die Länge (in Zeilen oder Wörtern) des einzelnen Sprechakts in Relation zu seinem Kontext (strukturelle Merkmale und Charakter) gesetzt, wie die folgende Abbildung 17 illustriert:

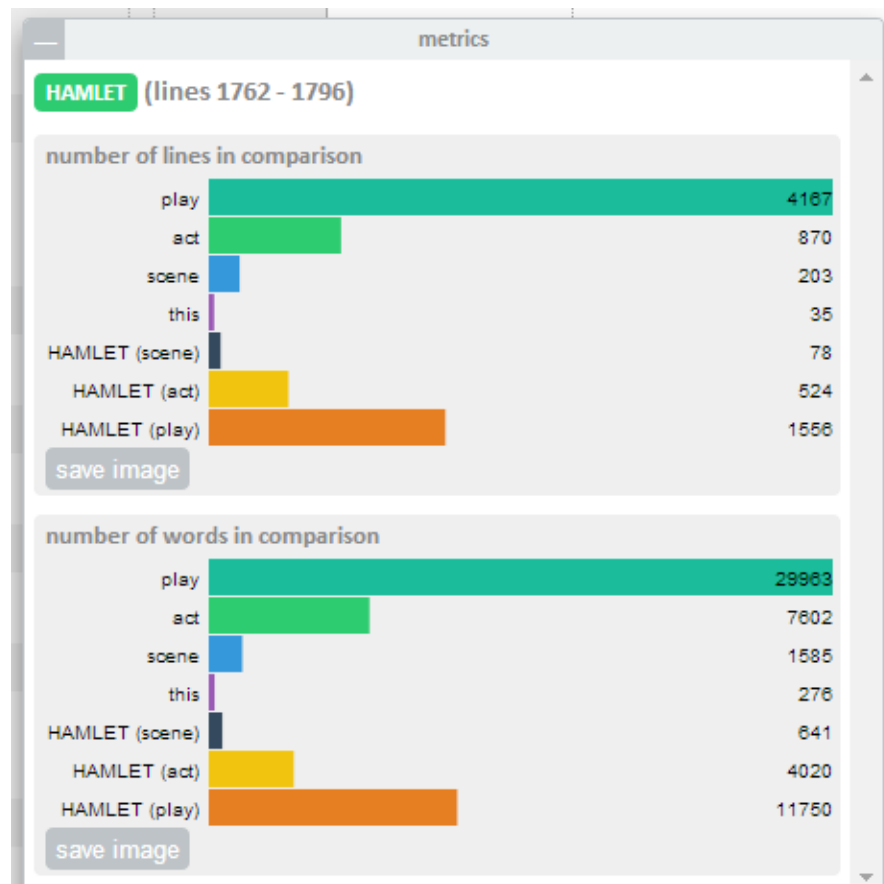


Abbildung 17: Länge von *Hamlets* Sprechakt „To be or not to be“ im Kontext

Der in Abbildung 17 gewählte Sprechakt ab Zeile 1762 ist 35 Zeilen bzw. 276 Wörter lang (mittlerer lila Balken „this“). Diese Zahlen werden in Relation zum Stück, umgebenden Akt und umgebender Szene dargestellt. Zudem werden sie im Verhältnis zu den Sprechanteilen des zugehörigen Charakters gezeigt: Etwa die Hälfte von *Hamlets* Anteil an der Szene macht dieser eine Sprechakt aus (Er spricht hier 35 Zeilen lang, sein Anteil in dieser Szene sind 78 Zeilen). Der aktuelle Sprechakt wurde in der Mitte platziert, um eine optische Unterscheidung der beiden Vergleichsdimensionen zu ermöglichen.

4.6.4 Bühnenanweisungen lesen

Der letzte Aspekt der Dramen, mit dem der Benutzer interagieren kann, sind die Bühnenanweisungen. Sie sind am oberen Rand direkt unterhalb der Strukturelemente angebracht. Im Unterschied zu den vorherigen Elementen bieten die Bühnenanweisungen bei Klick keine weiterführenden Informationen. Ihr Inhalt wird angezeigt, bewegt der Benutzer die Maus über die jeweilige Anweisung.

Die unterschiedlichen Typen sind durch Form und Farbe kodiert, die Legende links neben den Anweisungen zeigt sie dem Benutzer an. In der Annotation sind folgende

Typen vorgesehen: Betreten (**enter**) und Verlassen (**exit**) der Bühne durch einen oder mehrere Charaktere, allgemeines Geschehen wie Wetterphänomene oder Geräusche (**business**), der Ort an dem ein Charakter spricht (**location**), die Art wie dieser spricht (**delivery**), stummes Spiel (**dumbshow**) sowie Modifikationen einer Person (**modifier**; laut der Dokumentation von Folger (2013c) in der Regel Verkleidungen). Gleichzeitige, unterschiedliche Anweisungen können im Element `<stage type="mixed">` zusammengefasst sein. Die Unterscheidung nach Typen und Angabe betroffener Charaktere erleichtert sowohl den Zugriff auf einzelne Anweisungen, da diese aus dem Fließtext herausgebrochen worden sind, als auch ihre Anwendung z.B. durch Hobby-Regisseure. Im Sinne der automatisierten Auswertung der Stücke besonders interessant sind die in den Bühnenanweisungen enthaltenen Referenzen auf Personen. In welcher genauen Beziehung sie zur Anweisung stehen ist zwar nicht angegeben, doch kann diese Eigenschaft der Folger-Texte sinnvoll genutzt werden.

Ist folglich angegeben, welche Personen von der Anweisung betroffen sind, so werden diese hervorgehoben (die nicht betroffenen Charaktere werden ausgeblendet). Ein besonders geeignetes Beispiel sind Bühnenanweisungen wie „All but *Hamlet* exit“ (*Hamlet*, Akt 1, Szene 2, Bühnenanweisung zu Zeile 322). Zumindest im Text wird nicht auf den ersten Blick deutlich, wer gemeint ist. Abbildung 18 verdeutlicht, dass in diesem Fall *Getrude*, *Claudius*, *Laertes* und *Polonius* betroffen, also zu diesem Zeitpunkt außer *Hamlet* noch auf der Bühne, sind. Der weiße Pfeil markiert die Bühnenanweisung. Die Referenzen werden durch die schwarzen Pfeile angezeigt.

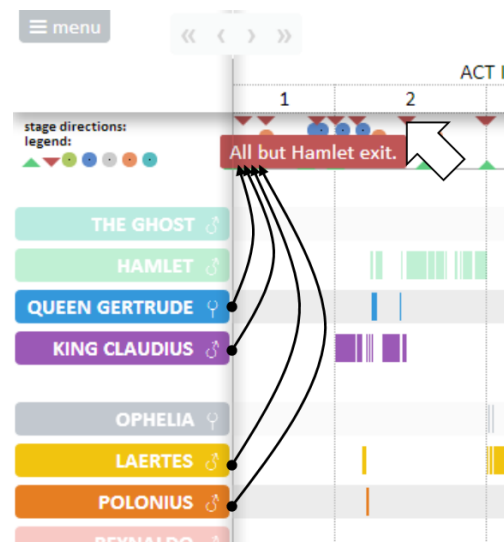


Abbildung 18: Bühnenanweisung mit Personenreferenzen

Die Anweisungen stehen ursprünglich an einer Stelle im Text, wo ihre Zuordnung zum Zeitpunkt der Handlung bzw. zu Personen, klar ist. Aufgrund der zweidimensionalen Darstellung ist es deshalb wichtig, betroffene Charaktere hervorzuheben, um dem Benutzer die Zuordnung zum Geschehen zu ermöglichen. Durch die Platzierung entlang der Zeitachse ist dies möglich. Des Weiteren kann das aktuelle Geschehen aus den in der Nähe befindlichen Sprechakten und ggf. Charakterreferenzen abgelesen werden.

4.6.5 Diagramme, Graphen und Rohdaten herunterladen

Unterhalb jeder Visualisierung im *metrics*-Fenster befindet sich ein Knopf mit der Aufschrift „save image“ zum Download der darüberstehenden Grafik auf den Computer des Nutzers. Diese Funktion ist für die gewinnbringende Nutzung von *To See or Not to See* erheblich: Inspiziert ein Benutzer ein beliebiges Element eines Stücks werden ihm zwar weitere Informationen angezeigt, er hat aber nicht die Möglichkeit diesen Stand festzuhalten. Durch den Download der Grafik wird diese Flüchtigkeit der Ergebnisse überwunden.

Die Grafiken liegen im Werkzeug als *Scalable Vector Graphics (SVG)* vor und müssen deshalb vor dem Download in ein Bildformat umgewandelt werden. Die Umwandlung wird serverseitig mithilfe der freien Linux-Grafikbibliothek *librsvg* vorgenommen (siehe Abschnitt 4.8.3).

Des Weiteren hat der Nutzer die Möglichkeit die Daten, auf denen sämtliche Visualisierungen beruhen, herunterzuladen und selbst weiterzuverarbeiten. Zu diesem Zweck befindet sich im Menü links oben der Unterpunkt „download source“. Für jedes

Stück ist eine *TSV*-Datei (*Tab Separated Values*) enthalten, in der die einzelnen Sprechakte und zugehörige Details (Sprecher, Länge, Inhalt u. Ä. durch Tabstopp getrennt) zeilenweise eingetragen sind. Sie können so z.B. als Datenbank in eine Software wie *R* geladen werden, um komplexere bzw. spezifischere Auswertungen vornehmen zu können.

4.7 To See or Not to See: Ein Rich-Prospect Browser?

In Teil 3 wurden *Rich-Prospect Interfaces* als theoretische Vorlage für *To See or Not to See* benannt und eingeführt. Ob das Werkzeug alle Faktoren des *Rich-Prospect Browsing* erfüllt, soll im Folgenden anhand der Funktionen geprüft werden. Nachfolgend stellt Tabelle 1 diese übersichtlich gegenüber. Die Faktoren richten sich nach der bereits weiter oben stehenden Liste von Ruecker et al. (2011, S. 3 f.).

Faktoren für <i>Rich-Prospect Browser</i>	Realisierung bei <i>To See or Not to See</i>
Alle <i>Informationseinheiten</i> sind in einer Übersicht sofort sichtbar	Das gesamte Interface, außer den schwebenden beiden Fenstern, zeigt <i>Abstraktionen</i> aller <i>Sprechakte</i> , <i>Charaktere</i> und <i>Bühnenanweisungen</i> (vgl. Abbildung 5)
Die <i>Informationseinheiten</i> können neu angeordnet werden	Nicht gegeben, weil die Anordnung dem Inhalt des Dramas folgt
Jede der <i>Informationseinheiten</i> bietet Zugang zu <i>weiterführenden Informationen</i>	Bei Interaktion werden <i>in den schwebenden Fenstern</i> zusätzlich kontextuelle Informationen angezeigt
Die Einheiten folgen einer <i>inhärenten Ordnung</i> und geben die verwendbaren <i>Werkzeuge</i> vor	Die Ordnung wird durch <i>Charaktere</i> und die <i>Zeitachse</i> bestimmt und ist jedem Sprechakt eingeschrieben. Die Interaktion durch <i>Klicken</i> ist das vorgegebene und einzige Werkzeug.
Der Benutzer kann, wenn möglich, zwischen verschiedenen <i>Darstellungsformen</i> der gleichen Datenkollektion <i>wählen</i>	Für die Übersicht <i>nicht erfüllt</i> , alternative Visualisierungen für die weiterführenden Informationen sind jedoch vorhanden
Die visuelle Ordnung der Übersicht soll dem Benutzer sogleich Aufschluss über ihre Bedeutung geben	Durch die Übersicht erfüllt: Die Achsen sind verständlich und leiten den Benutzer beim Lesen der nichtlinearen Textdarstellung an
Der Benutzer soll die einzelnen Einheiten markieren können, um sie im Falle einer Neuordnung wiederzuerkennen	Neuordnung nicht möglich; jedoch werden aktiv ausgewählte Sprechakte markiert (z.B. alle Sprechakte einer Szene)

Tabelle 1: Faktoren für *Rich-Prospect Browser* und ihre Erfüllung in *To See or Not to See*

Tabelle 1 veranschaulicht, dass *To See or Not to See* weitgehend als *Rich-Prospect Browser* eingestuft werden kann. Die wichtigste Eigenschaft, die Darstellung aller Informationseinheiten in einer Übersicht, ist weitgehend gegeben. Bei entsprechender Größe des Browserfensters sind die Sprechakte aller Personen sofort sichtbar. Andernfalls muss der Benutzer scrollen, um die weiter unten aufgeführten Charaktere zu sehen.

Zugriff auf weiterführende Informationen wird durch einen Klick gewährt und diese werden in den beiden schwebenden Fenstern angezeigt. Die Anordnung der Sprechakte erschließt sich sofort aus den beiden Achsen, an denen sie ausgerichtet sind. Position und Größe der Sprechakte können anhand dieses Koordinatensystems interpretiert werden: Wer wann (mit wem) und wie lange spricht, kann mit einem Blick wahrgenommen werden. Bei der visuellen Orientierung helfen vertikale Trennlinien für Akte und Szenen sowie die abwechselnde farbliche Hinterlegung der Zeilen für die Charaktere.

Die Informationseinheiten (Sprechakte, Personen, Bühnenanweisungen) neu anzuordnen ist nicht möglich und nicht gewollt: Eine Neukombination der Shakespeare-Texte ist nicht vorgesehen. Die optionale Forderung von Ruecker et al. (2011), dass der Benutzer verschiedene Darstellungsformen für die Gesamtkollektion zu Verfügung hat, wird nicht erfüllt. Der Inhalt der Stücke, also die Sprechakte, können nicht ausgetauscht werden. Möglich wäre hingegen als Richteinheit für die Zeitachse nicht die Zeilen des Stücks zu verwenden, sondern Worte. So würde die Darstellung etwas feiner. Technische Beschränkungen wie die Displayauflösung könnten hier jedoch noch schwerwiegender in Erscheinung treten, als bei der Orientierung an Zeilen.

Da keine Neuordnung der Informationseinheiten möglich ist, kann das letzte Kriterium konsequenterweise nicht eingehalten werden. Elemente (Sprechakte) werden nur dann markiert, wenn sie vom Benutzer zur weiteren Bearbeitung ausgewählt wurden. So werden alle Sprechakte einer Szene mit schwarzen Rändern markiert, wenn dieser angeklickt wird.

Für die Klassifizierung von *To See or Not to See* als *Rich-Prospect Browser* spricht des Weiteren, dass es keine Suchfunktion für Text oder andere Merkmale gibt. Die Interaktion ist damit auf das *Browsing* als Modalität – im Gegensatz zum Suchen – beschränkt. Insgesamt kann festgestellt werden, dass jene Kriterien, die nicht erfüllt wurden, der

Natur des Gegenstands geschuldet sind. Die wichtigsten Eigenschaften eines *Rich-Prospect Interface* trägt *To See or Not to See* somit in sich.

4.8 Technischer Aufbau

Bei *To See or Not To See* ist die Datengrundlage gleichzeitig Motivation und technische Bedingung. Das Datenformat legt dabei nicht nur die innere Struktur der einzelnen Dokumente fest, sondern gibt auch ihre Weiterverarbeitung grundlegend vor. Die in *TEI*-konformem *XML* vorliegenden Dokumente lassen sich mit verschiedenen Techniken auslesen bzw. auswerten. Mit einem *Parser* könnten so z.B. Detailinformationen (im einfachsten Fall etwa der Titel) ausgelesen werden. Verknüpft man derartige Anfragen an das Dokument programmatisch, ließe sich eine Umformung des Dokuments nach eigenen Vorstellungen, also in ein selbstdefiniertes Zielformat durchführen (etwa mehrere Tabellen einer Datenbank für die verschiedenen Aspekte der Dramen). Dies hat den Vorteil, dass auf spezifische Bedürfnisse der späteren Anwendung bis ins Detail eingegangen werden kann. Der Nachteil ist, dass ein solches Verfahren zeitaufwändiger ist, als bestehende Standards zu nutzen, welche der Struktur der Daten von vorne herein gerecht werden: Eine Transformation in ein einziges Zieldokument, bei der man selbst keine Überlegungen zum Ablauf oder zur Verschachtelung von Abfragen beim *Parsing* anstellen muss. Diese werden allgemein gültig, für beliebige *XML*-Dokumente von einer Transformationssoftware bereitgestellt. In diesem Fall bleibt dann lediglich das Zielformat in Form von Transformationsanweisungen selbst zu definieren, das im Gegensatz zur anderen Variante weniger Freiheiten bietet. In diesem Kontext bedeutet weniger Spezialisierung gleichermaßen mehr Allgemeingültigkeit. Zwar könnte mit der *Transformationsengine* ein beliebiges Format erschaffen werden, jedoch bietet die für *XML* entwickelte Markupsprache *XSL* (*Extensible Stylesheet Language*) zusammen mit einem *XSL*-Transformationsprozessors die Möglichkeit, Daten in *HTML* zu transformieren – ein anderes Format auf Basis der gleichen Meta-Markupsprache *XML*. Dies ist hier in zweierlei Hinsicht zielführend: Zum einen ist die visuelle Grundlage von Webanwendungen ein Gerüst aus *HTML*-Elementen, die bei der Transformation ohne einen weiteren Zwischenschritt erzeugt werden können. Zum anderen lassen sich diesen visuellen Elementen im gleichen Schritt statische Daten anhängen. An dieser Stelle werden visuelle Komponente und Daten eng aneinander gebunden. Im Vergleich zur Vorhaltung der Daten

in einer Datenbank, können sie damit direkt von Skripten oder Stilanweisungen angesprochen werden. Jedoch fehlt für eine weitere Verarbeitung (z.B. Datenauswahl für Graphen und Grafiken) eine mächtige Abfragesprache wie sie *SQL* bietet. Diese muss dann auf anderem Weg ergänzt werden (hier wird mit Hilfe von *jQuery* ein pseudo-*SQL*-Verhalten nachgebildet).

Besteht für eine Webanwendung lediglich das *HTML*-Gerüst, also die visuelle Komponente, kann wie bei anderer Software nicht von einer Applikation gesprochen werden. Ihr fehlt sowohl Funktionalität (z.B. Datensammlung s.o.) als auch Interaktivität. Diese werden durch die Integration von *JavaScripts* ermöglicht. Der Weg von den statischen Ausgangsdaten zur interaktiven Webanwendung ist in Abbildung 19 zusammengefasst.

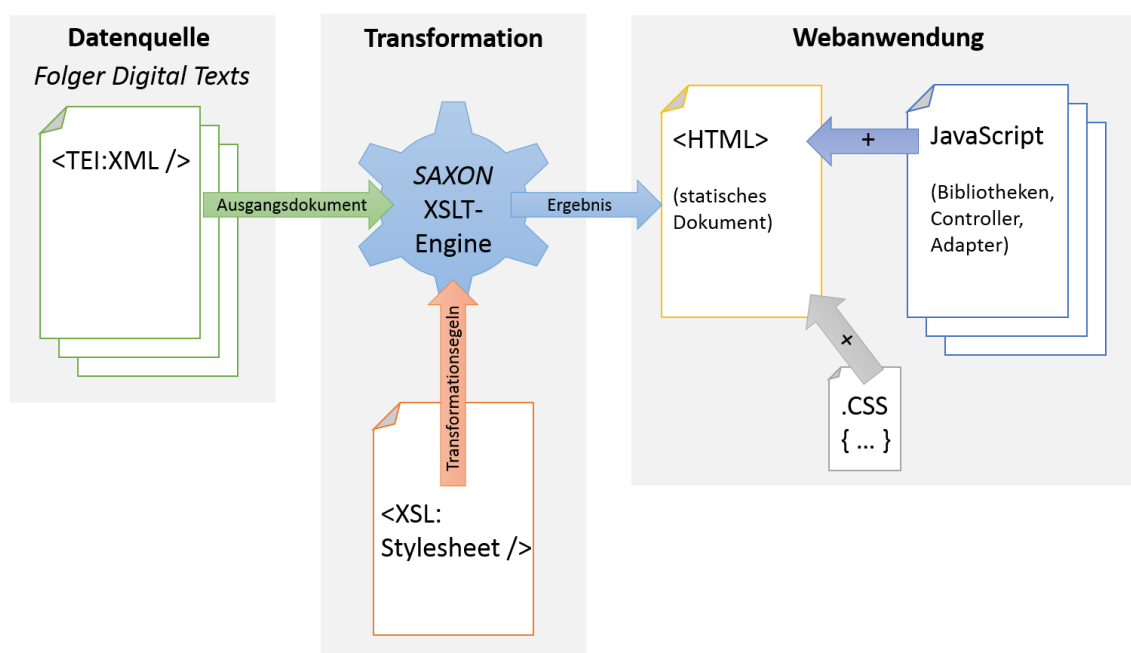


Abbildung 19: technischer Aufbau und Datenverarbeitungsprozess

Mithilfe der Transformationssoftware *Saxon*²⁸ werden alle Dramen auf Basis des gleichen XSL-Stylesheets transformiert. Das Ergebnis ist für jedes der Stücke eine *HTML*-Datei, in der sämtliche Aspekte in den entsprechenden Elementen samt Daten (z.B. Länge in Zeilen) enthalten sind. Durch Einbindung externer *JavaScripts* und Stilanwei-

²⁸ Verwendet wird der XSLT-Prozessor *Saxon-HE* von *Saxonica* (im Web unter <http://www.saxonica.com/welcome/welcome.xml>, letzter Zugriff am 25.4.2014) in der Version 9. Er liegt der Arbeit auf der CD bei.

sungen in *Cascading Style Sheets* (CSS), wird aus diesen statischen Dokumenten die interaktive Webanwendung. Details zum Transformationsvorgang werden in Abschnitt 4.8.2 aufgeführt.

In der Webanwendung selbst kommen verschiedene Techniken zum Einsatz, die durch Bibliotheken unterstützt werden. Abbildung 20 illustriert die einzelnen Komponenten und ihr Zusammenspiel.

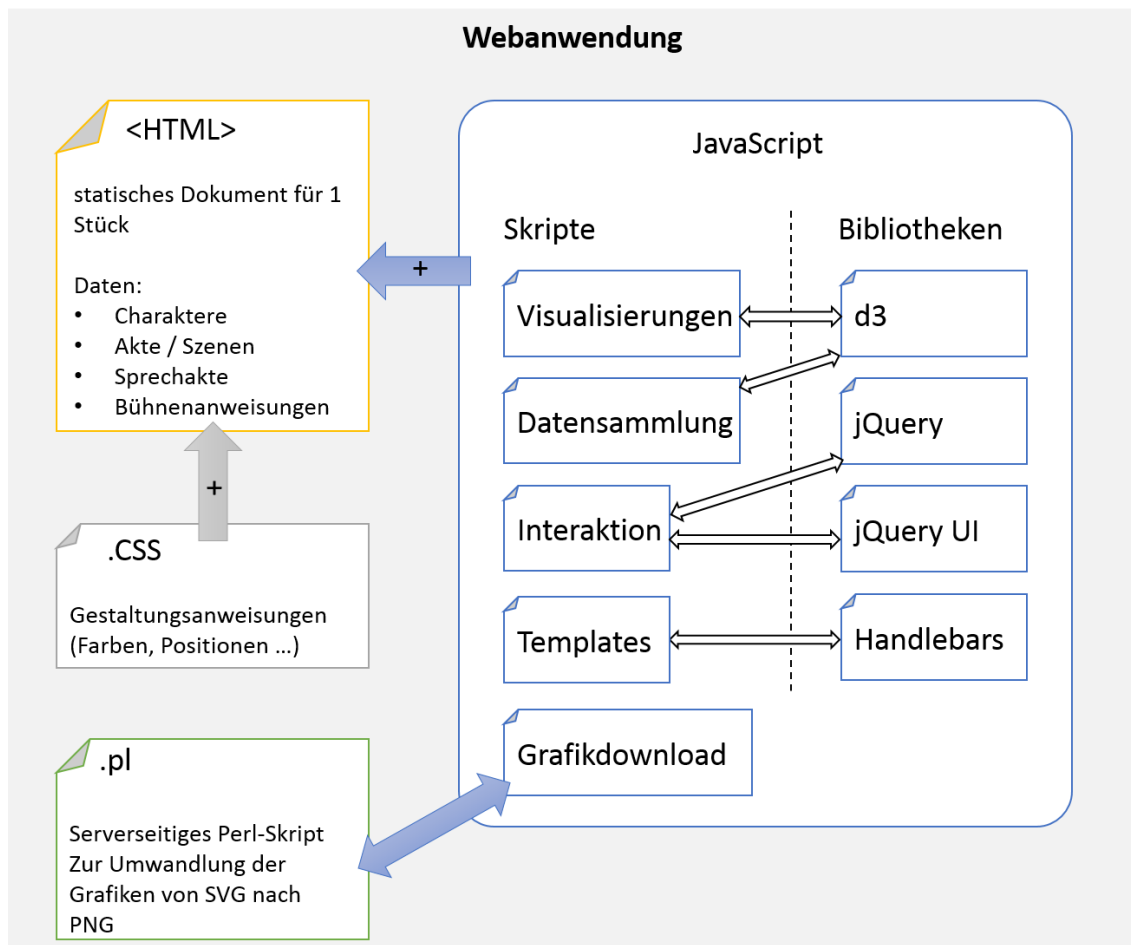


Abbildung 20: Komponenten der Webanwendung

Die Daten des Stücks werden statisch im *HTML*-Dokument vorgehalten, das gleichzeitig die Oberfläche der Anwendung bildet. Es enthält für jeden Charakter eine Zeile in der sich ein Feld für den Namen und all seine Sprechakte in Form von Blockelementen befinden. Jedes dieser Blockelemente ist dabei um Datenattribute erweitert (z.B. um den Wert `data-scene`, um anzuzeigen, in welcher Szene der Sprechakt zu finden ist). Insofern bilden alle diese Sprechaktelemente die Datenbasis für das gesamte Stück und können als Tabelle einer relationalen Datenbank verstanden werden: Jedes Element ist eine

Zeile. Die Spalten sind die Datenattribute. Lediglich der gesprochene Inhalt ist nicht als Datenelement sondern direkt in einem Kindelement als Absatz hinterlegt.

Alle diese Elemente und die restliche Oberfläche werden mithilfe von CSS angeordnet und Ihnen ihr Aussehen verliehen. So werden z.B. die Farben für die Charaktere alterniert oder die einzelnen Sprechakte an den richtigen Punkt auf der Zeitachse verschoben.

Das Dokument bindet weiterhin verschiedene *JavaScripts* ein, um Interaktion und Funktion des Werkzeugs zu ermöglichen. Darunter sind Standardbibliotheken der Webentwicklung. *jQuery*²⁹ ist eine der beliebtesten *JavaScript*-Bibliotheken, da sie den Zugriff auf bestimmte Elemente über komplexe Selektoren und viele Funktionen bietet, die häufig verwendet werden. So können auf einfache Weise Elemente, wie die Charakternamen ausgewählt und ihnen Ereignisse, wie Klicks, zugewiesen werden. Die Funktionalität von *jQuery* wird hauptsächlich für die Interaktion verwendet. Des Weiteren ermöglichen es die komplexen Zugriffsmöglichkeiten auf einzelne Elemente des *Document Object Model (DOM)*, Selektoren zu konstruieren, die ähnliche Aufgaben wie eine *Data Query Language (DQL)* im Umfeld von relationalen Datenbanken erfüllen. Dazu gehört die Auswahl bestimmter Elemente unter gegebenen Einschränkungen (z.B. alle Elemente, die zur Klasse der Sprechakte gehören und deren Akt-Attribut einen bestimmten Wert tragen, selektieren).

Früher Teil der Mutterbibliothek und heute ein eigenständiges Projekt ist *jQuery UI*³⁰, das häufig verwendete Funktionen, insbesondere für die Interaktion mit Webanwendungen, bereitstellt. Es nutzt intern *jQuery* und ist um reine Oberflächenfunktionen erweitert. Bei *To See or Not to See* kommt es zum Einsatz, um die schwebenden Fenster für die Betrachtung von Sprechakten und quantitativen Daten verschiebbar und in der Größe veränderbar zu machen.

Zur Erzeugung der Ansichten in den beiden schwebenden Fenstern wird *Handlebars*³¹ verwendet. Hierbei handelt es sich um eine *Templating Engine*. Sie ermöglicht die

²⁹ Im Web unter <http://jquery.com/>, letzter Zugriff am 25.4.2014. Wird in der Version 1.10.2 verwendet und direkt über das Google CDN bezogen.

³⁰ Im Web unter <http://jqueryui.com/>, letzter Zugriff am 25.4.2014. Wird in der Version 1.10.3 verwendet und ebenfalls über CDN bezogen.

³¹ Im Web unter <http://handlebarsjs.com/>, letzter Zugriff am 25.4.2014. Wird in der Version 1.3 verwendet und liegt dem Quellcode auf der CD bei.

statische Definition von einzelnen, beliebig komplexen Layoutvorlagen, die dann dynamisch bei Bedarf mit den entsprechenden Daten befüllt werden. Anschließend können diese an geeigneter Stelle ins bestehende Gesamlayout eingespeist werden. Für jeden Datentyp (z.B. Charakter) wird eine Ansicht definiert (z.B. alle Details zu einem Charakter, persönlicher Graph, Anzahl der Sprechakte usw.; s.o.).

Zur Generierung der Graphen und Grafiken wird *d3.js*³² (von engl. *data-driven documents*) benutzt. Es bringt ähnlich wie *jQuery* seine eigenen Selektierungsmechanismen mit, verfolgt damit aber andere Ziele. Dies sind die Berechnung von Datenvisualisierungen auf Basis beinahe beliebiger Datensätze und deren Anzeige als *SVG*. Das Endformat ist kein Bild im Sinne eines Dateiformats sondern Teil des *DOM* und somit wie alle anderen Elemente direkt manipulierbar. Um beliebige Daten in eine gewünschte Visualisierung zu überführen, ist es Aufgabe des Entwicklers einen Adapter zu implementieren, der beschreibt wie die Daten behandelt und dargestellt werden sollen (z.B. welcher Teil der Daten Werte bzw. Beschreibungen sind). Die Bibliothek bringt einige Visualisierungsformen (wie Tortendiagramme, Balkendiagramme und gewichtete Netze) sowie nützliche Funktionen (etwa die Ermittlung von Grenzwerten oder automatische Farbskalen) mit. Die Daten, die zur Visualisierung benötigt werden, können mithilfe der mächtigen Abfragemöglichkeiten von *jQuery* aus den relevanten Elementen ausgelesen und dann mit konventioneller *JavaScript*-Programmierung zusammengeführt werden. Diese Datenbündel müssen dem im Adapter festgelegten Muster folgen, um eine korrekte Visualisierung von *d3.js* berechnen lassen zu können.

4.8.1 Datengrundlage Folger Digital Texts

Die Basis für das Werkzeug bilden die bereits vorgestellten *Folger Digital Texts*. Dabei handelt es sich um eine digitale, annotierte Version der dem Hause eigenen englischsprachigen Edition (Folger Shakespeare Library, 2013a)³³. Im Unterschied zur gedruckten Variante enthält sie ausschließlich den Text des Stücks, ohne Anmerkungen, Hinweise auf Veränderungen im Vergleich zur Vorlage o. ä. (Folger Shakespeare Library, 2013b). Die Edition, auf der die digitalen Texte basieren, wurde zwischen 1994 und 2010

³² Im Web unter <http://d3js.org/>, letzter Zugriff am 25.4.2014. Wird in der Version 3 direkt über den Server von *d3* bezogen.

³³ Abgerufen von *About Us* am 28.4.2014 von <http://www.folgerdigitaltexts.org/>

von Barbara A. Mowat und Paul Werstine erarbeitet. Für Folger sind die *Digital Texts* die erste elektronische Veröffentlichung ihrer Werke.

Der entscheidende Unterschied im Vergleich zu anderen Versionen wie *Open Source Shakespeare* liegt in der Granularität der Annotation. Bei den *Folger Digital Texts* sind sämtliche Wörter, Zeichen und Referenzen zwischen den einzelnen Entitäten enthalten. Dagegen sind die Texte beim OSS nur grundlegend nach strukturellen (Akte, Szenen, Zeilen) und inhaltlichen (*Dramatis Personae*) Gesichtspunkten gegliedert in einer Datenbank gespeichert. Zwar enthält das Projekt auch eine Konkordanz, für alle Funktionen, die in *To See or Not to See* versammelt sind, würde diese Datenbasis nicht ausreichend spezifische Daten bereitstellen. Jedoch können Benutzer die Datenbank herunterladen, auf der die Webanwendung basiert. Ähnlich wie bei den Texten von *Folger* besteht also die Möglichkeit eigene Weiterentwicklungen vorzunehmen. Die entscheidende Rolle, welche Annotation bei der automatischen Verarbeitung und Aufbereitung von Texten einnimmt, wurde in Abschnitt 2 thematisiert und in Abschnitt 3 an verwandten Projekten exemplifiziert.

Zum Zeitpunkt der Veröffentlichung erster vorläufiger Ergebnisse dieses Projekts bei Wilhelm et al. (2013) standen die folgenden zwölf Werke bei den *Folger Digital Texts* zur Verfügung:

- *Hamlet*
- *Julius Caesar*
- *King Lear*
- *Macbeth*
- *Merchant of Venice*
- *A Midsummer Night's Dream*
- *Much Ado About Nothing*
- *Othello*
- *Romeo and Juliet*
- *Taming of the Shrew*
- *The Tempest*
- *Twelfth Night*

Im Oktober 2013 kamen *The Comedy of Errors* und *The Winter's Tale* hinzu. Die bestehenden Stücke wurden dabei zusätzlich kleineren Korrekturen unterzogen. Diese betrafen hauptsächlich einige der bisherigen Fehler in der Annotation. Sie hatten sich bei der Arbeit am Transformationsprozess bemerkbar gemacht. Nach einem kurzen Hinweis im Frühjahr 2013 an Michael Poston, der sich bei Folger für die Annotation verantwortlich zeichnet, versicherte dieser, bis Ende des Jahres viele der bekannten Fehler ausmerzen und weitere Stücke veröffentlichen zu wollen.³⁴ Dies war insbesondere mit Blick auf die Wiederverwendbarkeit der Transformationsanweisungen verheißungsvoll: Ob neu hinzukommende Texte problemlos mit der bestehenden Software umgewandelt werden können – also die von Moretti inspirierte Ausgangsthese gestützt werden kann – konnte somit noch vor Projektende weitgehend erfolgreich versucht werden.

Die Annotationsgrundlage für die *Folger Digital Texts* bildet das *Tagset*, welches von der *Text Encoding Initiative*³⁵ (TEI) für Dramen (und andere Texttypen) erarbeitet worden ist. Die Rolle von TEI für die literaturwissenschaftliche Forschung und der Einsatz des *Tagsets* wurden bereits weiter oben (Abschnitt 2 und 3) erörtert. Die Richtlinien zum Einsatz dieses de-facto Standards sind bei (The TEI Consortium, 2014) gesammelt. Mithilfe der auf XML-basierenden Markupsprache können die wesentlichen Eigenschaften von Theaterstücken erfasst werden. Neben Metainformationen wie Autor, Titel und Editoren sind strukturelle und inhaltliche Merkmale auszuzeichnen. Neben der einfachen Auszeichnung einzelner Elemente können zudem Referenzen zwischen ihnen hergestellt werden. So ist es z.B. möglich, für jeden Charakter der *Dramatis Personae* eine Referenz zu hinterlegen, die anschließend bei Bühnenanweisungen, Sprechakten u.Ä. zur Identifizierung dieser Person benutzt werden kann. Die Annotationstiefe ist bei der Anwendung des TEI-Standards beliebig. Sie geht bei Folger im Gegensatz zu vergleichbaren Projekten besonders weit: Die kleinste annotierte Einheit des Texts sind einzelne Wörter und Satzzeichen. Nachfolgend gibt Tabelle 2 einen Überblick über die von Folger verwendeten Elemente und Attribute des *TEI-Tagsets* und inwiefern diese bei *To See or Not to See* verwendet werden. In der Dokumentation der Folger Shakespeare Library (2013c) werden die Elemente und ihr Einsatz beschrieben.

³⁴ Erstkontakt mit Michael Poston per E-Mail (mposton@folger.edu) am 22.03.2013

³⁵ Im Web unter <http://www.tei-c.org/>, letzter Zugriff am 23.4.2014

Markup	Einsatz	Anwendung
<div1> , <div2>	Bei Textabschnitten, hier Akt und Szene; Enthalten entsprechende Kindelemente, hier Sprechakte und Bühnenanweisungen	Werden zur Anzeige der Zeitleiste ausgelesen und ihr Inhalt ausgezählt.
<lb> , <pb>	Zeilenumbrüche, Seitenumbrüche in der zugrundeliegenden gedruckten Ausgabe	Werden ignoriert, da sie für die Darstellung unerheblich sind.
<sp>	Innerhalb <div1> bzw. <div2> ; enthalten beliebig lange Sprechakte; tragen Attribute xml:id zur eindeutigen Nummerierung und tei:who zur Referenzierung des Sprechers	Jedes dieser Elemente wird später zu einem Sprechaktelement, das an entsprechender Stelle eingeordnet wird (auch auf Basis Attribute).
<speaker>	Innerhalb <sp> ; enthält noch einmal den Namen des Sprechers für die Darstellung	Bleibt ungenutzt, da die Angabe des Sprechers mittels des who -Attributs in <sp> vorgenommen wird.
<ab>	Innerhalb <sp> ; enthält den Text des Sprechakts	Wird als Selektierungspunkt zum Auslesen der Zeilen verwendet
<join>	Innerhalb von <ab> ; zeigt eine Zeile an: Nachfolgende (nicht innerhalb liegende!) Elemente gehören zu dieser Zeile; zu diesem Zweck mit tei:target -Attribut zur Referenzierung der Wörter und Satzzeichen	Zwischen jedem <join> -Element liegt eine Zeile des Texts, weshalb sie zur Bildung der Absätze innerhalb der Sprechaktelemente verwendet werden.
<w> , <pc> , <c>	Innerhalb von <ab> , <speaker> und <stage> (u.a.); zeigen Wörter, Satzzeichen und Leerzeichen an.	Werden zu Zeilen zusammengefasst und in dieser Granularität im Zielformat nicht weiter ausgezeichnet. Die Wortzählung wird auf Basis des <w> -Elements bei der Transformation vorgenommen.
<stage>	Beliebige Einordnung im XML-Baum; Enthalten Bühnenanweisungen und deren Typ im tei:type -Attribut	Werden direkt in die Bühnenanweisungselemente überführt (inkl. Datenattributen).
<listPerson>	Vor dem eigentlichen Text; enthalten zusammengehörige Personen (<person>); Können durch Einsatz	Werden zur Gruppierung der Einzelpersonen am linken Bildschirmrand verwendet.

	von <head> innerhalb benannt werden	
<person> , <personGrp>	Innerhalb von <listPerson> ; zeigt Einzelperson oder Personengruppen an; Enthalten Namen (<pers-Name>), Zusätze (<state>) und Geschlecht (<sex>)	Werden zur Erstellung der Charakterliste am linken Bildschirmrand verwendet. Die Kindelemente werden zudem ausgewertet.
<seg> , <foreign> , <q>	Innerhalb von <ab> , teils verschachtelt; <seg> enthalten Textsegmente der Typen letter , song , verse und poem (tei:type -Attribut), <foreign> in einer Fremdsprache (inkl. Angabe der Sprache durch ein Attribut) oder Zitate (<q>)	Werden benutzt um diese Abschnitte in den Sprechakten farblich zu kennzeichnen.

Tabelle 2: Bei den *Folger Digital Texts* eingesetztes TEI-Markup und Anwendung in *To See or Not to See*

Die Übersicht zeigt grundlegend die Struktur, in der die Elemente auftreten können. Hinzu kommen bei Elementen zu Personen, **<seg>** und **<stage>**, dass diese auch ineinander verschachtelt sein können. Hier kommt die rekursive Natur der *XSL*-Transformation vorteilhaft ins Spiel. So müssen Baumtiefen nicht etwa selbst bestimmt und traversiert werden. Unterordnungen, z.B. bei verschiedenen Bühnenanweisungen, die unter einer Anweisung vom Typ **<stage type="mixed">** gruppiert werden, werden hierdurch jedoch aufgelöst und gleichrangig wiedergegeben.

Die Auszeichnung von Wörtern und Satz- sowie Leerzeichen ist angegeben, da sie zu einzelnen Zeilen zusammengefasst werden. In den transformierten *HTML*-Dokumenten sind sie nicht mehr einzeln ausgezeichnet. Für die Auswertungen, die in *To See or Not to See* gemacht werden, ist dies kein Nachteil. Es sei erwähnt, dass die Zusammensetzung der Zeilen aufgrund ihrer Modellierung ein aufwändiger Vorgang ist. Die Zeilen werden nicht etwa als *Container*-Elemente annotiert, sondern als Einzelemente zwischen Gruppen von Wörtern und Satzzeichen platziert. Alle Wörter und Satzzeichen zwischen zwei **<join>**-Elementen gehören zu einer Zeile.

Elemente, die auf die gedruckte Ausgabe der Texte bezogen sind (Zeilen- und Seitenumbrüche) werden bei der Transformation ignoriert, da sie auf ein spezifisches Medium bezogen sind und so der Darstellung in nicht-linearer Form nicht dienlich sind.

Mit den genannten Elementen ist es (weitgehend) möglich, das Transformations-Stylesheet zu entwerfen, das schließlich zu der Übersichtsansicht von *To See or Not to See* führt. Weitergehende Funktionen, die z.B. Verbindungen zwischen Personen und anderen Elementen in Form von Graphen herstellen, basieren auf Attributen in den oben genannten Elementen.

Diese bereits genannten Personenreferenzen in Form von Attributen sind eine Besonderheit der Folger-Texte, die unterschiedliche Aspekte des Stücks mit den zugehörigen Charakteren verbinden. Die technische Grundlage dafür wird durch Angabe der Personenschlüssel in den entsprechenden Elementen geschaffen, die der Ausschnitt aus dem *TEI*-Dokument in Abbildung 21 zeigt. Abgebildet ist die zweite Bühnenanweisung zur Zeile 322 (Attribut `xml:id`) vom Typ `exit` (markiert das Verlassen der Bühne), die schon in Abschnitt 4.6.4 als Beispiel herangezogen wurde.

```

6615 <stage xml:id="stg-0322.1b" type="exit" ana="#CLAUDIUS-HAM
      #GERTRUDE #POLONIUS #LAERTES">
6616   <w xml:id="w0050040" n="SD 1.2.132.1">All</w>
6617   <c xml:id="c0050050" n="SD 1.2.132.1"> </c>
6618   <w xml:id="w0050060" n="SD 1.2.132.1">but</w>
6619   <c xml:id="c0050070" n="SD 1.2.132.1"> </c>
6620   <w xml:id="w0050080" n="SD 1.2.132.1">Hamlet</w>
6621   <c xml:id="c0050090" n="SD 1.2.132.1"> </c>
6622   <w xml:id="w0050100" n="SD 1.2.132.1">exit</w>
6623   <pc xml:id="p0050110" n="SD 1.2.132.1">.</pc>
6624 </stage>

```

Abbildung 21: Bühnenanweisung mit Personenreferenzen im *TEI*-Ausgangsdokument

Im `<stage>`-Element, welches einzelne Bühnenanweisungen anzeigt, ist das Attribut `ana` enthalten. Darin werden die betroffenen Charaktere über ihre Referenzen (`#` und `Personenname`) identifiziert. Diese Daten werden ins Zieldokument übernommen und anschließend durch eine Komponente des *JavaScripts* ausgewertet. Dies wird notwendig, da sich komplexere Strukturen, wie etwa eine Liste, nicht in den Datenattributen der *HTML*-Elemente speichern lassen.

4.8.2 Transformationsprozess

XML-basierte Strukturen wie sie hier vorliegen können per XSL-Transformation in ein neues Format überführt werden. Hierzu wird ein in XSL abgefasstes Dokument erstellt, das Regeln festlegt nach denen die Transformation durch den XSLT-Prozessor ablaufen soll. Diese Regeln beziehen sich dabei nicht auf konkrete XML-Dokumente sondern de-

ren Schema. In der Praxis bedeutet dies, dass mit einem Regelsatz beliebig viele, gleichartige Dokumente transformiert werden können. Das Ergebnis folgt dabei ebenfalls immer dem Schema, das in den Transformationsregeln als Ausgabeformat festgeschrieben worden ist. Für dieses Projekt heißt dies konkret, dass alle aktuellen Stücke mit dem bestehenden Regelsatz umgewandelt werden können. Dies ist insbesondere in Bezug auf den hier formulierten Anspruch hin relevant, beliebige Dramen, die *TEI*-konform ausgezeichnet sind, transformieren zu können.

Ein weiterer Vorteil wird deutlich, wenn neue Texte zur Kollektion hinzukommen: Während der Entwicklung des Werkzeugs wurden von Folger zwei weitere Stücke veröffentlicht (s.o.). Die Transformation mit dem bestehenden Regelsatz war weitgehend erfolgreich. Es waren demnach keine neuen Aspekte enthalten, für deren Modellierung noch keine entsprechenden Transformationsregeln vorliegen. Sollte dies aber bei beliebigen zukünftigen Texten der Fall sein, ist eine Erweiterung kaum problematisch. Das liegt an der Natur des Transformationsvorgangs. Die Regeln werden „nach Auftreten“ definiert. Tritt also ein bestimmtes Phänomen im Text auf, wird es entweder von bestehenden Regeln verarbeitet, d.h. in eine spezifisch festgelegte Ausgabeform gebracht, oder ignoriert bzw. so wie es auftritt wiedergegeben. Letzteres hängt vom Aufbau des *Stylesheets* ab.

Wird also bei der Fehlerüberprüfung des neu transformierten Stücks ein neues Phänomen³⁶ gesichtet, so wird für dieses eine Regel definiert. Diese wird nur dann angewendet, wenn es im ursprünglichen Text auftritt. Die bereits erfolgreich erfassten Stücke bleiben davon unberührt. Um also beinahe beliebige Dramen mit dem Werkzeug verarbeiten zu können, müssten sämtliche Dramenvarianten und ihre Spezialphänomene, die in *TEI* modellierbar sind, ermittelt werden. Dies ist praktisch für die Dramen Shakespeares mit Sicherheit machbar, für beliebige Stücke aus unterschiedlichen Epochen aber kaum vorstellbar. Immerhin handelt es sich hierbei um ein Problem, das auch die Definition des *TEI*-Standards betrifft.

³⁶ Bei einfachen Phänomenen wie Liedern, die in Sprechakten vorkommen, ist die Integration direkt durchführbar. Hierzu wird die Markierung in den Regelsatz für die Transformation von Sprechakten aufgenommen. Komplexere Phänomene, wie Charaktere, die in Verkleidung auftreten, erfordern unabhängig von ihrer Modellierung absehbar mehr Aufwand.

Für *Too See or Not To See* folgt daraus, dass die berechnete Prognose gestellt werden kann, alle von Folger noch zu veröffentlichenden Stücke auch in kürzester Zeit integrieren zu können. Die relative Homogenität, die durch die Beschränkung auf einen Autor gegeben ist, ist hierfür entscheidend. Wie bei jedem Werkzeug ist dies eine Abwägung zwischen Spezifität und Vielseitigkeit.

Das Ergebnis der Umwandlung des *XML*-Baums in einen *HTML*-Baum macht deutlich, dass es sich lediglich um eine Reorganisation bekannter Informationen handelt: Daten werden für den neuen Verwendungszweck verknüpft, geordnet und aufbereitet – Es ergibt sich ganz im Sinne der Ausgangsidee ein „neuer Blick“ auf die Daten. Dadurch allein ergibt sich noch kein informationeller Mehrwert. Deshalb werden während der Transformation Daten zur Länge von Abschnitten und Sprechakten u.Ä. erhoben und zusätzlich in den zugehörigen Elementen hinterlegt. Durch die Interaktion mit dem Text in seiner reorganisierten Form werden so neue Erkenntnisse ermöglicht.

Das Ziel der Anwendung ist ein datenbasierter Blick auf die Werke Shakespeares. Insofern ist es konsequent, die Texte in eine Form zu überführen, die strukturiert(er) und automatisch verarbeitbar(er) – in Bezug auf die Anwendung – ist. Es wurde bereits weiter oben angesprochen, dass die einzelnen Aspekte des Stücks im Ausgabeformat als Relationen dargestellt werden. Um in der Metapher des relationalen Datenbankmodells zu bleiben, ist jedes im Enddokument enthaltene Element ein Tupel. Dieses Tupel gruppiert sich durch seine Modellierung (und eine Klassenbezeichnung) zur zugehörigen Relation. So sind alle Charaktere, Sprechakte und Bühnenanweisungen nach dem gleichen Muster hinterlegt. Für einen Charakter sieht das Schema wie folgt aus (Tabelle 3), Zeilen und Spalten sind transponiert:

Klasse	Charakter	Charakter
Charakter	Hamlet	Ophelia
Geschlecht	männlich	weiblich
Beschreibung	Prince of Denmark, son of the late King Hamlet	
Stirbt in Zeile...	4114	
Stirbt nicht vor Zeile...		3123
Stirbt nicht nach Zeile...		3366

Tabelle 3: Schema der Relation für Charaktere mit Beispieldaten

Das Beispiel zeigt zwei Elemente der Klasse **Charakter**, die durch weitere Informationen genauer bezeichnet werden. Hier zwar als Relation dargestellt, sind aber nicht alle dieser Daten „flach“ mit dem Element verknüpft. Die zusätzliche Beschreibung einer Person (im Beispiel *Hamlet* als *Prince of Denmark*) liegen im HTML-Baum in einem eigenen Kindelement. Durch die Mächtigkeit der Selektierungsmöglichkeiten mittels *jQuery* ist dies keine Einschränkung, sondern kann gleichzeitig zu Darstellungszwecken genutzt werden: Bei *Mouseover* wird die zusätzliche Beschreibung angezeigt.

Die gleiche Vorgehensweise wird für Sprechakte und Bühnenanweisungen angewendet. So ergeben sich am Ende vier verschiedene Relationen, die durch *HTML*-Elemente repräsentiert werden: Charaktere, Sprechakte, Akte und Szenen. Aus diesen Elementen und den hinterlegten Daten werden sämtliche Auswertungen und Visualisierungen generiert. Diese statische Vorhaltung der Daten zu einem Stück im *HTML*-Gerüst ist hier im Vergleich zur Verwendung einer Datenbank aus mehreren Gründen ausreichend und adäquat. Die Daten ändern sich nur, wenn eine neue Transformation durchgeführt wird. Diese läuft dann ab, wenn eine neue Version der Texte von Folger veröffentlicht wird. Die Datenbasis als solche ist wenig dynamisch, weshalb der Einsatz einer Datenbank nicht notwendig wird. Neben Performance-Überlegungen spielt auch der Vorteil eine Rolle, das visuelle Grundgerüst gleichzeitig als Datenquelle zu nutzen. Das *HTML*-Markup bringt diese Verschränkung von Daten und Darstellung mit und wird hier ausgenutzt. Die durch die Transformation entstandene, angereicherte, statische Datenbasis wird durch die Einbindung von Skripten zu einer interaktiven Webanwendung, wie der nächste Abschnitt zeigt.

4.8.3 Interaktive Webanwendung

Mit der Einbindung diverser Bibliotheken und eigenem *JavaScript*-Code werden die Interaktionsmöglichkeiten von *To See or Not to See* realisiert. Weiter oben und in Abbildung 20 wird ein Überblick über die verwendeten Bibliotheken und ihre Einsatzgebiete gegeben. An dieser Stelle soll im Detail geklärt werden, welche Funktionen unter Zuhilfenahme welcher Bibliothek implementiert sind.

Der Regler, der an der Zeitachse zur Navigation durch das Stück verwendet werden kann, stützt sich auf die Selektierungs- und Sortierungsmechanismen, die in *jQuery* integriert sind. Der restliche Standard-*JavaScript*-Code ist so modelliert, dass der Regler flexibel manipulierbar ist und somit auf die verschiedensten Nutzeraktionen reagiert.

Sämtliche dieser Interaktionen (Tastatureingaben, Klicks und Mouseover) werden mithilfe von *jQuery* überwacht. Nur dort, wo mit Bordmitteln von CSS das Anzeigen von Informationen bei Mouseover möglich war (z.B. beim Menü), wurde dementsprechend auf *JavaScripts* verzichtet.

Die beiden schwebenden Fenster *speech* und *metrics* stellen neben der Überblicksansicht die wichtigste Komponente dar. Sie werden mithilfe von CSS über dem restlichen Inhalt angezeigt und durch zwei in *jQuery UI* enthaltene Funktionen in ihrer Größe veränderbar und verschiebbar. Grundsätzlich ist dies keine komplexe technische Hürde, jedoch bringt die Bibliothek Optimierungen und Kompatibilität mit, die ansonsten kaum erreichbar wäre. Dass die Fenster verschiebbar sind und versteckt werden können, ist für die Übersichtlichkeit und Benutzbarkeit wichtig. Als Vorbild wurden hier Programme wie *Adobe Photoshop* genommen, bei denen die gleiche Metapher verwendet wird, um die Benutzeroberfläche flexibel und nutzerfreundlich zu gestalten. In der ersten Entwicklungsstufe waren die Fenster nicht integriert und die Benutzung stark eingeschränkt. Mit den beiden Fenstern besteht heute die Möglichkeit auch komplexere Informationen darzustellen, ohne das Interface zu überladen.

Der Inhalt der Fenster ist je nach gewähltem Aspekt des Stücks unterschiedlich. Bei jedem Klick werden diese nach Layoutvorlagen neu generiert und an der entsprechenden Stelle in das *DOM* integriert. Wie oben erwähnt, wird hierfür eine weitere Bibliothek verwendet. *Handlebars* ermöglicht es, komplexe *HTML*-Gerüste als Vorlagen zu definieren und an bestimmten Stellen flexible Informationen anzuzeigen. Des Weiteren stehen programmierähnliche Steuerungsmechanismen wie *if-else*-Verzweigungen oder Schleifen zur Verfügung. Die Vorlagen werden zur Laufzeit ausgewertet und mit den Informationen bestückt, die der *Templating Engine* übergeben werden. Ähnlich wie bei *d3.js* wird zuvor durch den Programmierer eine Schnittstelle für jedes Layout definiert. Für die Sprechakte, die im Fenster *speech* angezeigt werden, ist so z.B. eine Vorlage definiert, die den Namen des Charakters farbig hinterlegt anzeigt und darunter den Text samt Zeilenzahlen anzeigt. *Handlebars* erwartet entsprechend Daten zu Namen, Farbcode, Text und Zeilenzahlen. Je nachdem wie viele Sprechakte angezeigt werden sollen (z.B. alle eines bestimmten Charakters), werden von der Bibliothek die passende Zahl Layout berechnet und schließlich im Fenster angezeigt. Die gleiche Vorgehensweise gilt für die anderen Informationstypen im Fenster *metrics*.

Unterhalb dieser Informationen befinden sich die verschiedenen eingesetzten Diagramme und Graphen. Sie werden durch *d3.js* berechnet. Die verschiedenen Adapter, welche das Datenformat und die Visualisierungsform festlegen und zusammenbringen, nutzen zugleich Funktionen der Bibliothek wie auch Standard *JavaScript*. Die Daten zur Visualisierung werden mittels komplexer *jQuery*-Selektoren aus den *HTML*-Elementen ausgelesen und algorithmisch so vorverarbeitet, dass sie durch die Bibliothek visualisiert werden können. Weiter oben wurde bereits die Analogie zu Abfragen in relationalen Datenbanken erwähnt. Hier werden die Vorteile dieser Vorgehensweise für die Vorbereitung der Visualisierungen genutzt.

Die relevanten Werte für die einzelnen Darstellungen werden erst zur Laufzeit aus den statischen Daten ausgelesen. Grundsätzlich wäre es möglich, für eine unveränderliche Datenbasis, wie sie hier vorliegt, sämtliche möglichen Visualisierungen bzw. ihre zugrundeliegenden Datensätze im Voraus zu berechnen. Da der Berechnungsaufwand relativ gering ist (es werden z.B. keine Massendaten verarbeitet), ist dies jedoch nicht zwingend erforderlich.

Da die Visualisierungen nicht schon vorberechnet werden, können Sie nicht direkt vom Benutzer als Grafik gespeichert werden. Durch *d3.js* generierte Diagramme und Graphen werden als *SVG* in den *DOM* eingespeist. Zwar können sie aus diesem Grund von *CSS*-Regeln optisch verändert werden und sind mittels *jQuery* ansprechbar, was die Interaktion (z.B. das Verschieben der Knoten in den Personennetzen) mit ihnen ermöglicht. Für den Benutzer ist es jedoch z.B. nur durch das Anfertigen eines Screenshots möglich, die Grafiken weiterzuverwenden. Abhilfe wird durch eine Umwandlung der *SVG*-Daten in *PNG*-Bilder geschaffen. Ermöglicht wird diese durch die serverseitige, freie Grafikbibliothek *librsvg*. Sie enthält ein Programm, das für diesen Zweck geschaffen worden ist. Es nimmt vorverarbeitete *SVG*-Ausschnitte des *DOM* als Parameter entgegen und berechnet daraus die Rastergrafik. Um es anzusprechen ist eine Kommunikation zwischen dem Server und der Webanwendung auf dem Client zu ermöglichen. Hierzu wurde auf dem Server ein *Perl*-Skript platziert, das vom *JavaScript* aus angesprochen werden kann. Sobald das Bild aus den übergebenen Daten berechnet worden ist, beginnt im Browser des Nutzers automatisch der Download der Datei. Das Skript wurde

ursprünglich vom *GitHub*-Nutzer A. Gordon unter *BSD*-Lizenz bereitgestellt.³⁷ Die vorgesehenen Wahlmöglichkeiten hinsichtlich des Ausgabeformats wurden für die Anwendung in *To See or Not to See* gestrichen (nur noch *PNG* als Zielformat) und eine dynamische Namensgebung für die Bilder integriert, so dass diese auch nach dem Download zugeordnet werden können.

³⁷ Im Web unter https://github.com/agordon/d3export_demo, letzter Zugriff am 22.4.2014. An dieser Stelle ergeht der ausdrückliche Dank an den Autor. Die modifizierte Variante des Skripts läuft auf dem Server der Medieninformatik.

4.9 Projektevolution

Die Rolle der evolutionären Weiterentwicklung von Projekten als Folge von Erfolgen und Misserfolgen wurde weiter oben bereits als charakteristisch für die *Digital Humanities* identifiziert. *To See or Not to See* bildet hierbei keine Ausnahme. Deshalb soll der folgende Abschnitt einen Einblick in den Entwicklungsverlauf der Applikation geben – insbesondere, um den hier vorgestellten Stand von den bereits veröffentlichten Ergebnissen bei Wilhelm et al. (2013) abzugrenzen.

4.9.1 Bis zur Erstveröffentlichung im Mai 2013

Im Rahmen der Vorbereitung eines Seminars zu den *Digital Humanities* im Winter 2013 wurde Manuel Burghardt auf die *Folger Digital Texts* und die daraus resultierenden Möglichkeiten aufmerksam. Gemeinsam wurde die erste Version von *To See or Not to See* als Seminarprojekt auf den Weg gebracht, welche schließlich im Mai 2013 auf der Konferenz *Kultur und Informatik* in Berlin einem wissenschaftlichen Publikum präsentiert werden konnte (Wilhelm et al., 2013).

Die Hauptaufgabe in dieser ersten Phase war es, ein Verständnis für die Textmodellierung im *TEI*-Format zu gewinnen und daraus Möglichkeiten abzuleiten, wie der Text auf alternative Weise dargestellt werden kann. So entstand die Überblicksansicht, die dem Werkzeug heute zugrunde liegt auf Basis einer *XSL*-Transformation von *TEI* nach *HTML*. Die Orientierung des Texts entlang einer Zeit- und Personenachse und die Reduktion der Sprechakte auf Markierungen innerhalb des so entstandenen Koordinatensystems wurden in dieser Phase umgesetzt. In diesem Zuge legten wir uns auf die Zeilen des Stücks als Abstraktion der Zeit fest.

Nach Entwicklung der Übersichtsansicht konnte die Integration weiterer Ideen aus dem *Rich-Prospect Browsing* geplant werden. In einem ersten Schritt wurden grundlegende, zu den strukturellen und inhaltlichen Elementen gehörende Informationen an die jeweilige Repräsentation angehängt und bei *Mouseover* sichtbar gemacht. Hierzu gehörten beinahe ausschließlich Informationen, die in den Ausgangstexten bereits hinterlegt waren. Bei Sprechakten waren dies etwa deren Inhalt oder bei Charakteren neben deren Namen das Geschlecht, ein Namenszusatz und der Todeszeitpunkt. Quantitative Aspekte spielten nur in rudimentärer Form eine Rolle: Die Länge von Sprechakten, Akten, Szenen in Wörtern und Zeilen; Die Anzahl enthaltener Sprechakte und Bühnenanweisungen in den einzelnen Abschnitten. Die Visualisierung derartiger Daten wurde

zum Ende dieser Phase angedacht und auf experimentelle Art umgesetzt. Für Akte und Szenen wurde ein Balkendiagramm generiert, das die Anzahl enthaltener Wörter, Zeilen und Bühnenweisungen in Relation zum Stück (und dem umgebenden Akt) setzte. Hierin sind die späteren Ideen zur Datenvisualisierung begründet.

Neben der Möglichkeit, mittels Mouseover die einzelnen Aspekte des Stücks zu untersuchen, war die Navigation durch das Stück anhand der Sprechakte mithilfe eines Schiebereglers möglich. Zum Zeitpunkt der ersten Veröffentlichung war dieser nur eingeschränkt funktional. Für diesen Regler und die Visualisierung quantitativer Aspekte der Stücke wurde ein komplexeres Datenmodell notwendig, das in der zweiten Phase umgesetzt worden ist.

4.9.2 Auswertung quantitativer Daten und Umstellung auf neues Datenmodell

Diese zweite Phase des Projekts baut auf neuer Inspiration und Kritik aus der Konferenz in Berlin auf. Es wurden neue Anforderungen definiert, die, mit Blick auf diese Arbeit, eine Konzentration auf wesentliche Aspekte und eine Verbesserung der Nutzerinteraktion darstellten. Zu diesen zentralen Aspekten gehört vor allem die Auswertung quantitativer Daten und deren Visualisierung.

Um diese umsetzen zu können, musste die bisherige, einfache Umformung der Daten vom einen ins andere Format in eine Transformation in ein Datenmodell umgestaltet werden. Zwar konnte dies unter Beibehaltung einiger Teile des bestehenden Transformationsvorgangs geschehen, die Anpassungen am Frontend waren jedoch von größerem Umfang. Hier wurde beinahe der gesamte Quellcode neu- bzw. umgeschrieben. Die äußere Erscheinung des Werkzeugs hat sich dabei nur geringfügig verändert, wie illustriert. Aus der Transformation der einen Darstellungsform (*TEI*) in eine andere Darstellungsform (*HTML*) wurde eine Transformation in eine neue Darstellungsform (*HTML* mit Datenattributen), die gleichzeitig als Datenmodell und –quelle dient³⁸.

³⁸ Die Nutzung von *HTML*-Elementen als quasi „Datenbank“ und die Implikationen für die Trennung von Inhalt, Funktion und Darstellung werden in Abschnitt 4.8 diskutiert.

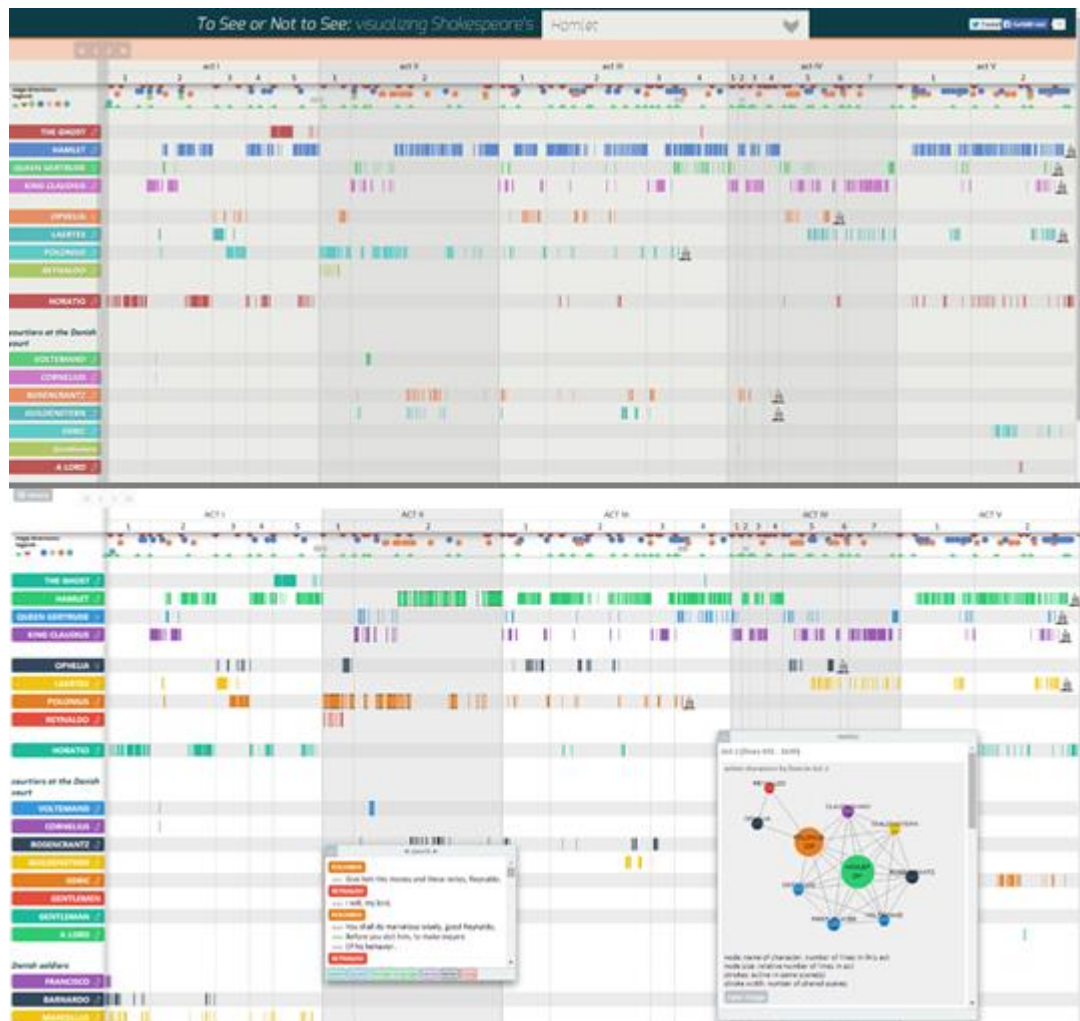


Abbildung 22: Vergleich der ersten (oben) und zweiten (unten) Version von *To See or Not to See*

Im gleichen Zug wurde die Funktionalität des Schiebereglers auf der Zeitachse optimiert. Durch eine engere Bindung an das Datenmodell besteht nun die Möglichkeit seine Position immer dann neu zu bestimmen, wenn der Benutzer mit einem Sprechakt interagiert oder auf andere Art durch das Stück navigiert.

Hauptaufgabe dieser Entwicklungsphase war jedoch die Implementierung der Datenvisualisierungen. Hierzu wurde zuerst eine Einarbeitung in die *d3.js*-Bibliothek vorgenommen. Im nächsten Schritt wurden die Datenschnittstellen für die unterschiedlichen Visualisierungsformen definiert. Mithilfe komplexer *jQuery*-basierter Abfragen auf die *HTML*-Datenbasis werden die Daten für die jeweiligen Diagramme oder Netze gesammelt. Zum aktuellen Zeitpunkt werden die grundlegenden Anforderungen an die Grafiken erfüllt, jedoch blieben notwendige zusätzliche Funktionen zu implementieren.

Dazu gehören sich nicht überlappende Beschriftungen bei allen Grafiken sowie Achsenbeschriftungen bei allen Diagrammen.

Um die Grafiken und andere Informationen in den beiden schwebenden Fenstern anzeigen zu können, welche ebenfalls erst in dieser Entwicklungsstufe zur Verbesserung der Benutzbarkeit eingefügt wurden, müssen die Daten hierfür nicht nur gesammelt und aufbereitet werden. Für die unterschiedlichen untersuchbaren Aspekte eines Stücks gibt es jeweils eine spezifische Ansicht in den beiden Fenstern. Sie werden bei jeder Nutzeraktion nach einem festgelegten Muster (*Templates*) aufgebaut. Für die Erzeugung der Ansichten wurde das *Templating System Handlebars* ins Projekt integriert. Tabelle 4 fasst diesen Abschnitt zusammen und gibt noch einmal einen Überblick darüber, in welcher Entwicklungsphase welche Funktionen implementiert wurden.

Funktion	1. Phase	2. Phase
Übersichtsansicht	weitgehend in heutiger Form vorhanden	kleinere visuelle Verbesserungen und Markierung aktiver Elemente; Die Elemente werden mit Daten angereichert und so zur "Datenbank" (Einführung des neuen Datenmodells)
Weiterführende Informationen ...	bei <i>Mouseover</i> des Elements werden diese in direkter Nähe angezeigt ...	Einführung der schwebenden Fenster, die die Informationen bei Klick anzeigen; Integration von <i>Handlebars</i> zur Erstellung der spezifischen Ansichten für alle Inhaltselemente
... zu Personen	Name, Geschlecht, Zusatztext, Todeszeitpunkt in der Übersicht	zusätzlich: Anzeige aller Sprechakte, Personengraph, Auswertungen zu Sprechakten in den versch. Abschnitten
... zu Abschnitten	Länge in Zeilen und Wörtern, Anzahl der enthaltenen Bühnenanweisungen, Liste aktiver Personen	Inhalte werden in Relation zu ihrem Kontext gesetzt, Ausblenden inaktiver Personen bei <i>Mouseover</i> , Personengraphen und Diagramme zu quantitativen Aspekten
... zu Sprechakten	Text, Länge in Zeilen und Wörtern, Anzeige, ob Spezialabschnitte (wie Gesang) enthalten sind	zusätzlich: Inhalt und quantitative Daten werden in Verhältnis zu ihrem Kontext gesetzt und visualisiert
... zu Bühnenanweisungen	Inhalt	zusätzlich: Hervorhebung betroffener Personen (bei <i>Mouseover</i>); keine weitere Auswertung in den Fenstern
Schieberegler	vorhanden, aber fehlerhaft	funktional
Visualisierungen	Zwei einfache Balkendiagramme für die Länge (in Wörtern und Zeilen) von Akten und Szenen im Verhältnis zum Stück (bzw. Akt) Erstellung für Testzwecke ohne <i>d3.js</i>	Einführung von <i>d3.js</i> zur Erstellung der Personengraphen, Balkendiagramme und Ringdiagramme für alle Aspekte.
Downloads		Daten- und Grafikdownload

Tabelle 4: Übersicht Entwicklungsphasen und Komponenten von *To See or Not to See*

5 Diskussion

5.1 To See or Not to See: born digital?

Beim Gegenstand des Interesses der vorgestellten Anwendung handelt es sich grundsätzlich um eine analoge – sogar für die mündliche Wiedergabe gedachte – Quelle. Digitale Ausgaben Shakespeares oder anderer bekannter Texte gab es schon vor der Veröffentlichung durch Folger. Ins Rollen brachte das Projekt jedoch nicht etwa die digitale Verfügbarkeit des Texts an sich, sondern seine spezielle Form.

Annotierte Texte bieten an sich ein Mehr an Information, das von außen kommt und letztlich der annotierenden Person unterliegt (Garside et al., 1997, S. 4). Der Detaillierungsgrad, wie ihn Folger anbietet, führt zu einer so grundlegenden Veränderung der Spielregeln im Umgang mit den Daten, dass die Texte von der *digitalisierten* zur *digitalen* Ressource werden. Sie bieten bisher kaum vorhandene Möglichkeiten, den Text alternativ darzustellen und zu analysieren. Schon allein aus diesem Grund bleibt *To See or Not to See* nicht im *Humanities Computing* verhaftet, sondern gehört zu den modernen *Digital Humanities* (Burdick et al., 2012, S. 11).

Das Produkt dieses Projekts ist, neben der vorliegenden Arbeit und der vorläufigen Vorstellung auf der Konferenz *Kultur und Informatik*, eine Webanwendung. Mit der freien Verfügbarkeit des Werkzeugs über das Internet wird es sowohl Wissenschaftlern wie Laien zugänglich. Die Möglichkeit, auf die Basisdaten zuzugreifen, macht es zu einer offenen Plattform, die eine Weiterverwendung und Dialog ermöglicht. Diese Plattform basiert dabei auf interdisziplinärer Arbeit. Die Verwendung der Daten Dritter, wie sie Folger und jetzt *To See or Not to See* ermöglichen, ist laut Ross (2014) im Sinne Morettis.

Die Zuordnung dieses Vorhabens zu den *Digital Humanities* ist somit klar. Inwiefern sich auch die Forderungen des *Distant Reading* verwirklichen lassen, soll im folgenden Abschnitt geklärt werden.

5.2 Eine Anwendung des Distant Reading?

Die Oberfläche von *To See or Not to See* zeigt gleichzeitig immer nur eines der Dramen. Ein Vergleich mehrerer Stücke ist Prinzipiell durch Aufruf dementsprechend vieler Instanzen der Anwendung möglich. Das Werkzeug unterstützt den Benutzer insofern beim *Distant Reading* der Theaterstücke, indem es vom eigentlichen Inhalt abstrahiert

Daten darüber anzeigt. Eines der zentralen Motive der Methode ist die Visualisierung ausgewählter Merkmale. Hierfür bietet das Werkzeug verschiedene Anknüpfungspunkte. Dabei ist eines der entscheidenden Merkmale, dass Ergebnisse gespeichert werden können. Insofern bringt es die Möglichkeit einer datengetriebenen, visuellen Analyse (z.B. aller Personennetze aller ersten Akte) auf den Weg: Die Auswertung der Grafiken ist im Anschluss Aufgabe des Benutzers.

Neben dem rein datenorientierten Blick ist zudem die Möglichkeit zum Close Reading eines Textes gegeben. Diese widerspricht laut Ross (2014) nicht den Prinzipien des *Distant Reading*. Sie sieht die Kombination beider Modi als gewinnbringend an:

[...] moving back and forth between the microscopy of close reading and the wide-angle lens of distant reading would enrich both methods, creating a dual perspective that boasts both specificity and significance. (S. 4)

Clement (2008) ignoriert den Anspruch des *Distant Reading*, Erkenntnisse über mehrere Texte zu gewinnen, und konzentriert sich auf die Arbeit an *einem* literarischen Text mithilfe datengetriebener Visualisierungen. Auf diese Weise legt sie die Wiederholungsmuster in Steins *The Making of Americans* offen, die ansonsten, so die Autorin, im aus der Nähe notorisch unlesbaren Text verborgen bleiben. Mit *Too See or Not to See* wird ein ähnlicher Weg eingeschlagen, indem in einer Anwendungssitzung typischerweise ein Text bearbeitet wird, mit dem Unterschied, dass Ergebnisse zu Einzeltexten gespeichert und so später weiterverwendet werden können.

Der Herangehensweise bei beiden Projekten liegt ein für digitale Werkzeuge zentrales Problem zugrunde. Mit der Bearbeitung einer bestimmten Textsorte, dem Einsatz einer bestimmten Technik oder der Implementierung spezieller Funktionen wird implizit eine Abwägung zwischen Spezifität und Vielseitigkeit getroffen. An den in Abschnitt 3 vorgestellten Projekten lässt sich diese Herausforderung exemplarisch ablesen: Während einige Werkzeuge mit (beinahe) beliebigen Texten arbeiten (*Voyant Tools*, *Mandala Browser*), verwenden andere (*WordSeer*, *SET*) XML-formatierte Daten. Die vorliegende Arbeit hat dabei gezeigt, dass mit zunehmend detailreicher Annotation komplexere automatische Verfahren möglich werden bzw. ihre Erstellung vereinfacht wird.

Für den Anspruch des *Distant Reading*, im Optimalfall alle jemals publizierten Texte als potentielle Datenquelle nutzen zu können, gibt es demnach drastische praktische

Einschränkungen. Während aus offensichtlichen Gründen der Machbarkeit mittelfristig nicht mit der Annotation eines Weltarchivs zu rechnen ist, zeigt das Beispiel von Clement, dass bestimmte Texte einer Spezialanwendung mit eingeschränktem Nutzungskontext erfordern. Um mit Moretti zu argumentieren, ist die Konzentration auf einen Text ohnehin zu vernachlässigen. Spezialanwendungen können – auf Basis welcher Anforderungen dies auch sein mag – auch für beliebige große Textkorpora notwendig sein. Insofern kann ein Werkzeug nie den globalen bzw. allen Ansprüchen des *Distant Reading* genügen. Vielmehr trägt jede in diesem Sinne erstellte Anwendung zur Bereicherung der Wissenschaft bei, indem neue Perspektiven aufgezeigt werden, die zu neuen Erkenntnissen führen – oder auch nicht.

5.3 Bekannte technische Probleme

Obwohl die funktionalen Anforderungen an die Anwendung klar festgelegt und über einen adäquaten Zeitraum erfüllbar waren, bleiben technische Fehler, die es in Zukunft zu beheben gilt. Bei einer komplexen Anwendung, wie sie hier entstanden ist, können die Gründe in den einzelnen Teilen oder im Zusammenspiel der eingesetzten Komponenten liegen. Durch den Einsatz von Fremdsoftware (Bibliotheken wie *jQuery* und *d3.js*) und -daten kann nicht in jedem Fall eine unmittelbare Lösung herbeigeführt werden.

Die Qualität der Transformation hängt einerseits von den Transformationsanweisungen ab. Andererseits ist die Fehlerfreiheit der Datenquelle für das Endergebnis entscheidend.³⁹ Da die Anweisungen strikt ausgeführt werden, führen selbst kleinere Fehler im Ausgangstext in der Konsequenz zu fehlerhaften Daten, Visualisierungen und, unter Umständen, Interpretationen. Trotz großer Sorgfalt auf Seiten von Folger sind derartige Datenprobleme bei der Annotation zu erwarten und hier auch deutlich geworden. Wie oben bereits erwähnt, zeigte sich Folger nicht nur in diesem Zusammenhang kooperativ. Neben der Beseitigung von Fehlern wurden auch neue und verbesserte Annotationsmittel in Aussicht gestellt. Für *To See or Not to See* ergeben sich daraus aller Voraussicht nach neue Möglichkeiten. Schließlich wirkt sich die Qualität der aktuellen Transformation auch auf die zum Download angebotenen Grunddaten aus, da diese auch auf dem

³⁹ Crane (2006) hat die Fehlerfreiheit digitaler Quellen mit Blick auf wachsende Kollektionen problematisiert, die sich hier praktisch niederschlägt (siehe Abschnitt 2.1)

Transformationsergebnis basieren. Es ist deshalb davon auszugehen, dass diese konsequenterweise ebenso nicht fehlerfrei sind.

Auf Markupebene sind die Stücke größtenteils homogen. Spezifische Einzelheiten einiger Werke hingegen benötigen eine Modellierung, die in dieser Form nur bei diesem Stück Anwendung findet. Um diese individuellen Phänomene abdecken zu können, müssen XSL-Transformationsregeln ergänzt werden. Durch die rekursive Anwendung dieser Regeln ist dies theoretisch problemlos machbar, sofern die Spezifika der Texte bekannt sind. Im Umfeld der Dramen Shakespeares sind die Durchsicht der einzelnen Stücke und die Anpassung der Transformation an deren Erfordernisse vertretbar und machbar. Sieht man sich in diesem Zusammenhang mit einem größeren Korpus konfrontiert, müssen Abwägungen zur Genauigkeit getroffen werden. Eines dieser Spezifika ist das Stück im Stück (*The Murder of Gonzago*) bei *Hamlet*, das nicht gesondert als solches ausgewiesen wird.

Die Visualisierungen auf Basis von *d3.js* sind zum aktuellen Zeitpunkt zwar informativ, aber z.T. schwer lesbar. Dazu gehören überlappende Labels bei Ringdiagrammen und Personengraphen. Des Weiteren könnten sich bei Ringdiagrammen Legenden als nützlich erweisen und Skalen die Lesbarkeit von Balkendiagrammen erhöhen. Dies sind kleinere Anpassungen, die aufgrund geringerer Priorität zu diesem Zeitpunkt noch ausstehend sind.

Der Download der Grafiken birgt bei den Graphen das Problem, dass im heruntergeladenen Bild die Kanten zwischen den einzelnen Knoten nicht sichtbar sind. Das Problem liegt im Zusammenspiel der Konvertierungssoftware in *librsvg* und den vom *JavaScript* gelieferten Daten. Die dynamisch positionierten Kanten werden in der Konvertierung nicht mit verarbeitet. Die Behebung dieses Fehlers erfordert einen Zwischenschritt, bei dem die einzelnen Kanten in eine statische Form gebracht werden, bevor die Umwandlung in das Bildformat vorgenommen wird. Dieses Problem hat für die Nutzbarkeit direkte Folgen und hat deshalb für die nähere Zukunft hohe Priorität.

Insgesamt ist bei so komplexem Material, wie es die Dramen Shakespeares darstellen, mittelfristig mit kleineren Fehlern im Detail zu rechnen, die wie eingangs beschrieben, unterschiedlichste Ursachen haben können. Es wird darauf gehofft, dass Benutzer, denen diese auffallen, sich an den Autor wenden, damit diese Fehler beseitigt werden können.

5.4 Ausblick

Die Idee, dass beliebige Shakespeare-Dramen mithilfe von *To See or Not to See* visualisiert werden können ist aufgegangen. Die zwischenzeitliche Veröffentlichung zweier neuer Stücke wurde als Gelegenheit genutzt, dieser Anforderung nachzugehen. Qualitativ ist die Transformation der neuen Stücke auf gleich hohem Niveau, wie die vorhandenen Texte. Für die zukünftige Veröffentlichung weiterer Texte von Folger ist also davon auszugehen, dass diese ebenso beinahe problemlos transformiert werden können. Allerdings bleibt *To See or Not to See* spezifisch auf das *TEI-Markup* von Folger beschränkt. Größere Flexibilität könnte durch eine Erstellung einer normalisierten Datenbank erreicht werden. Dazu bräuchte es spezialisierte Transformationsanweisungen für die jeweilige Modellierung der Texte aus anderen Quellen, die zeitaufwändig und somit eher im Rahmen eines größeren Projekts zu verwirklichen wäre.

Bisher konnten mit *To See or Not to See* keine Fallstudien durchgeführt werden, die seine Nützlichkeit im Forschungsprozess unter Beweis stellen könnten. Es bleibt abzuwarten, ob Wissenschaftler oder Laien Interesse am Werkzeug finden und welche Erkenntnisse oder welchen Lerneffekt sie damit erzielen können. Der Funktionsumfang ist hierfür zwar grundlegend gegeben, jedoch blieben Möglichkeiten flexiblere Visualisierungen zu ermöglichen, bei denen der Benutzer auch auf die Verwendung bestimmter Daten Einfluss nehmen kann. Bei den Charakternetzen zu Akten könnte dem Benutzer die Möglichkeit geboten werden, interaktiv weitere Akte hinzuschalten bzw. ausblenden. Denkbar wäre zudem eine Erweiterung, die es ermöglicht, einzelne Personen auszuwählen und deren spezifisches Subnetz anzuzeigen bzw. hervorzuheben. Es wurde weiter oben bereits problematisiert, dass eine derartige Anwendung im Spannungsfeld von Spezialisierung und allgemeiner Anwendbarkeit existiert. Insofern ist die Nützlichkeit für Wissenschaftler von der Spezifität ihres Forschungsinteresses abhängig, da *To See or Not to See* nicht zur Beantwortung einer spezifischen Forschungsfrage entworfen worden ist.

Literaturverzeichnis

- Berry, D. M. (2011). The Computational Turn: Thinking About the Digital Humanities. *Culture Machine*(12). Abgerufen am 28. April 2014 von <http://www.culturemachine.net/index.php/cm/article/view/440/470>
- British National Corpus. (2014). *How the BNC was created*. Abgerufen am 16. April 2014 von About the British National Corpus: <http://www.natcorp.ox.ac.uk/corpus/index.xml?ID=creation>
- British Library. (kein Datum). *Partners*. Abgerufen am 11. April 2014 von Shakespeare in quarto: <http://www.bl.uk/treasures/shakespeare/partners.html>
- Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., & Schnapp, J. (2012). *Digital Humanities*. Cambridge, Massachusetts, USA; London, England: The MIT Press.
- Busa, R. A. (1980). The Annals of Humanities Computing: The Index Thomisticus. *Computers and the Humanities*(14), S. 83-90.
- Busa, R. A. (2004). Foreword: Perspectives on the Digital Humanities. In S. Schreibman, R. Siemens, & J. Unsworth (Hrsg.), *A Companion to Digital Humanities* (S. XVI). Oxford: Blackwell.
- Clement, T. E. (2008). 'A Thing not beginning and not ending': using digital tools to distant-read Gertrude Stein's *The Making of Americans*. *Literary and Linguistic Computing*, 23(3), S. 361-381.
- Crane, G. (März 2006). What Do You Do with a Million Books? *D-Lib Magazine*, 12(3). Abgerufen am 28. April 2014 von <http://www.dlib.org/dlib/march06/crane/03crane.html>
- DHd. (2014). *DHD Unconference*. Abgerufen am 21. April 2014 von Digital Humanities im deutschsprachigen Raum: <https://dig-hum.de/dhd-unconference>
- Folger Shakespeare Library. (2013a). Abgerufen am 2. April 2014 von Folger Digital Texts: <http://www.folgerdigitaltexts.org/>
- Folger Shakespeare Library. (2013b). *FAQs about the Folger Library Shakespeare Editions*. Abgerufen am 2. April 2014 von Folger Shakespeare Library: <http://www.folger.edu/Content/About-Us/Publications/Folger-Editions/FAQs-about-the-Folger-Library-Shakespeare-Editions.cfm>
- Folger Shakespeare Library. (2013c). *Folger Digital Texts Documentation*. Abgerufen am 22. April 2014 von Folger Digital Texts: http://www.folgerdigitaltexts.org/fdt_documentation.pdf

- Garside, R., Leech, G. N., & McEnery, T. (1997). *Corpus annotation. Linguistic information from computer text corpora*. London; New York: Longman.
- Johnson, E. M. (2005). *How Moby Shakespeare Took Over the Internet*. Abgerufen am 11. April 2014 von Open Source Shakespeare:
http://www.opensourceshakespeare.org/info/moby_shakespeare.php
- Keim, D. A., & Oelke, D. (2007). Literature Fingerprinting: A New Method for Visual Literary Analysis. *IEEE Symposium on Visual Analytics Science and Technology* (S. 115-122). Washington, DC, USA: IEEE Computer Society.
- Leipzig Corpora Collection. (1. April 2014). *About the Leipzig Corpora Collection*. Abgerufen am 16. April 2014 von Leipzig Corpora Collection:
<http://asvdoku.informatik.uni-leipzig.de/corpora/>
- McLuhan, M. (2001). *Understanding Media*. London: Routledge.
- Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., & Aiden, E. L. (2011). Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, 331(6014), S. 176–182.
- Moretti, F. (2007). *Graphs, maps, trees. Abstract models for a literary history*. London: Verso.
- Moretti, F. (2013). *Distant reading*. London, New York: Verso.
- Mullin, M. (2003). Shakespeare on the Web. In L. Davis (Hrsg.), *Shakespeare Matters: History, Teaching, Performance* (S. 119-137). Newark: University of Delaware Press.
- Muralidharan, A., & Hearst, M. A. (2013). Supporting exploratory text analysis in literature study. *Literary and Linguistic Computing*, 28(2), S. 283-295.
- Ramsay, S. (2011). *Reading machines: toward an algorithmic criticism*. Urbana, Chicago and Springfield: University of Illinois Press.
- Roberts-Smith, J., DeSouza-Coelho, S., Dobson, T. M., Gabriele, S., Rodriguez-Arenas, O., Ruecker, S., Sinclair, S., Akong, A., Bouchard, M., Hong, M., Jakacki, D., Lam, D., Kovacs, A., Northam, L., & So, D. (2013). Visualizing Theatrical Text: From Watching the Script to the Simulated Environment for Theatre (SET). *DHQ*, 7(3).
- Rockwell, G., Sinclair, S., Ruecker, S., & Organisciak, P. (20. Dezember 2010). Ubiquitous Text Analysis. *Poetess Archive Journal*, 2(1).
- Ross, S. (2014). In Praise of Overstating the Case: A review of Franco Moretti, *Distant Reading* (London: Verso, 2013). *Digital Humanities Quarterly*, 8(1).

- Ruecker, S., Radzikowska, M., & Sinclair, S. (2011). *Visual Interface Design for Digital Cultural Heritage: A Guide to Rich-prospect Browsing*. Farnham, Surrey, England; Burlington, VT: Ashgate.
- Schreibman, S., Siemens, R., & Unsworth, J. (2004). The Digital Humanities and Humanities Computing: An Introduction. In S. Schreibman, R. Siemens, & J. Unsworth (Hrsg.), *A Companion to Digital Humanities* (S. XXIII). Oxford: Blackwell.
- Serlen, R. (2010). The Distant Future? Reading Franco Moretti. *Literature Compass*, 7(3), S. 214-225.
- Sinclair, S., Ruecker, S., Gabriele, S., Patey, M., Gooding, M., Vitas, C., & Bajer, B. (2011). Meditating on a Mandala in Class: Studying Shakespeare's Plays with a Visual Exploration Tool for XML Texts. *Media : Culture : Pedagogy*, 15(1).
Abgerufen am 28. April 2014 von
http://mcp.educ.ubc.ca/v15n01BornDigital_Article01_Sinclair_Ruecker_Gabriele_Patey_Gooding_Vitas_Bajer
- Stanford Literary Lab. (kein Datum). *About*. Abgerufen am 3. April 2014 von
<http://litlab.stanford.edu/>
- Svensson, P. (2009). Humanities Computing as Digital Humanities. *Digital Humanities Quarterly*, 3(3).
- Svensson, P. (2010). The Landscape of Digital Humanities. *Digital Humanities Quarterly*, 4(1).
- The TEI Consortium. (20. Januar 2014). *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Abgerufen am 23. April 2014 von Text Encoding Initiative:
<http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>
- Wilhelm, T., Burghardt, M., & Wolff, C. (2013). "To See or Not to See". An Interactive Tool for the Visualization and Analysis of Shakespeare Plays. In R. Franken-Wendelstorf, E. Lindinger, & J. Sieck (Hrsg.), *Kultur und Informatik: Visual Worlds & Interactive Spaces* (S. 175-185). Glückstadt: Verlag Werner Hülsbusch.

Erklärung

Ich habe die Arbeit selbständig verfasst, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und bisher keiner anderen Prüfungsbehörde vorgelegt.

Unterschrift